IBM research

PROJECT LIGHTNING

tenth quarterly progress report

technical appendixes

Volume II

PROJECT LIGHTNING
Tenth Quarterly Progress Report
1 March 1961 to 31 May 1961
(Contract NObsr 77508, Task 4)
Technical Appendixes

Volume II

International Business Machines Corporation
Research Center
Yorktown Heights, New York, 31 July 1961

CONTENTS

Technical Appendixes

APPENDIX I

A Memorandum on Results of Tin and Indium Cryotrons

A. E. Brennemann

ABSTRACT

This memorandum presents general results on the characteristics of tin and indium cryotrons evaporated in conventional evaporators with no special techniques such as substrate heating or prenucleation.

Indium from the conventional system is compared with films from more elaborate systems. Indium from the conventional system compared favorably with that produced in an ultra-high vacuum system.

Reproducibility results for four in-line cryotrons on one substrate indicate that the cryotrons cannot be interconnected with a sufficient margin of safety on the bias or overdrive on the control when maximum operating speed is desired.

The critical currents for a number of tin and indium films are plotted as a function of the ratio of film thickness to penetration depth.

The gain characteristics of unity crossing crossed-film cryotrons are discussed.

INTRODUCTION

The results given below are taken from data on tin and indium cryotrons that were made in conventional evaporators without substrate heating or prenucleation. One purpose of this paper is to compare the sharpness of the magnetic transition and the magnetic hysteresis of these materials with the same characteristics of those materials prepared in more elaborate systems. In particular, the general characteristics of indium are compared with indium prepared in the systems of A. Toxen[1] and H. Caswell.[2]

Reproducibility of both the material characteristics and the cryotron circuit characteristics are necessary in order to interconnect cryotrons either within one substrate or on separate substrates. Results on reproducibility are given below for indium cryotrons and the margin of safety in circuit operation is illustrated.

Results on critical currents of indium and tin films can be related to calculated results using the London field and current distribution. The experimental results are expressed in terms of βt, the thickness-to-penetration depth ratio, and can be useful in certain cryotron circuit designs even though incremental gain for the crossed-film[3] devices has not been thoroughly investigated.

The results given below are typical of the device characteristics from the conventional evaporator. It should be stated, however, that both tin and indium with desirable characteristics were prepared using the special techniques in the conventional evaporators. The special techniques could not be used in the device production system conveniently.

MAGNETIC PROPERTIES OF INDIUM DEVICES

A general comparison of the magnetic field transitions of tin and indium prepared in a conventional evaporator is given in the ninth quarterly Lightning report. Indium films, as obtained from a number of different evaporator systems, are compared in Table I. It may be seen that films produced in a conventional system compare favorably with H. Caswell's films produced in very high vacuum. A. Toxen's films, produced in a conventional vacuum and then mechanically trimmed, showed the more desirable characteristics, but the technique for producing these films would not be compatible with device production techniques.

No films were made by the more sophisticated techniques of prenucleation and substrate heating for comparison in this study.

Table I - Results for Indium Films
at Reduced Temperature Values of $t^2 < 0.8$

	I Conventional device films	II H. Caswell's films	III A. Toxen's films
Magnetic transition widths, $\Delta H/H$	3-7%	3-4%	1-2%
Magnetic hysteresis $\Delta H_y/H$	10%	10%	5%
Substrate condition	room temperature	room temperature	nitrogen cooled
Film edge	untrimmed	untrimmed	trimmed

The resistance ratios of films from all three systems are comparable. The inference from Table I is that indium films from the conventional evaporator, with uncontrolled conditions, closely approach the others in characteristics and are to some extent suitable for device usage. It should be possible to produce films more nearly like those shown in column III by greater control of the production techniques. The critical temperatures of the indium devices from conventional evaporators were reproducible within a range of 3.385 to 3.415°K. Tin produced in the same evaporators had critical temperatures in a range of 3.73 to 3.90°K.

REPRODUCIBILITY OF CRYOTRONS

The characteristics of four indium in-line cryotrons across a substrate are given in Table II and the corresponding gain curves and resistance transitions are shown in Fig. 1a and 1b for a reduced temperature of $T = 0.896\ T_c$.

Table II - Reproducibility Results on Indium Cryotrons[a]

	Cryotron 1	Cryotron 2	Cryotron 3	Cryotron 4
Thickness	6800 Å	6950 Å	6620 Å	6900 Å
Critical field, H_c	68 oe	65 oe	64 oe	64 oe
Magnetic transition widths, $\Delta H/H$	5.9%	3.1%	3.1%	4.7%
Critical gate current	450 ma	420 ma	410 ma	410 ma
Critical control	1350 ma	1420 ma	1310ma	1400 ma
Control current transition width, $\Delta I_c/I_c$	13.7%	7%	10.7%	7.1%

[a] $T = 0.896 T_c$. Control width and gate width = 0.009 inch.

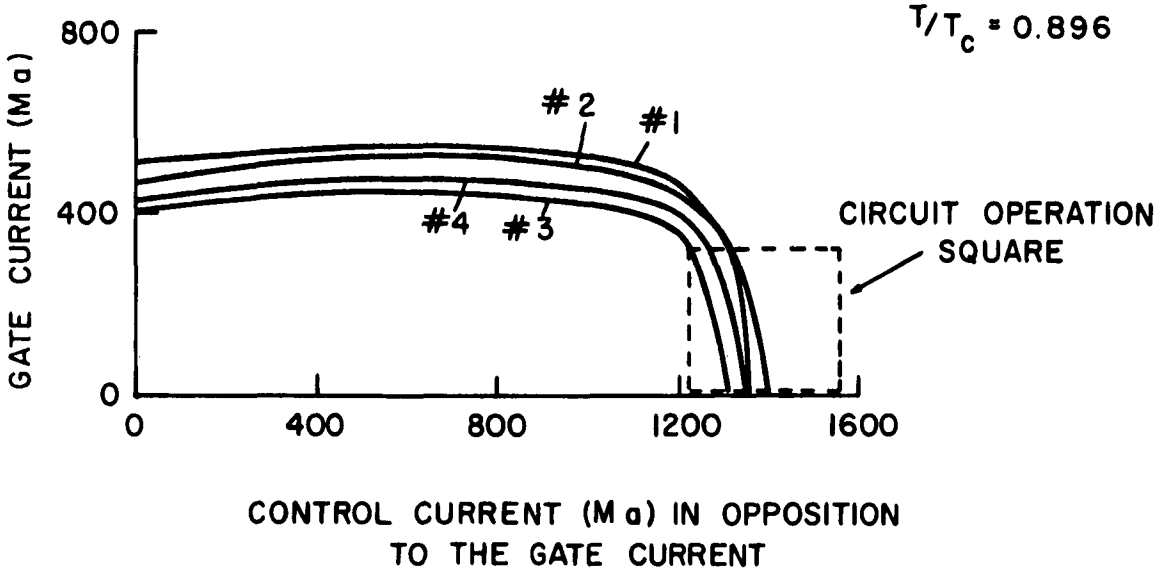


Figure 1a - In-line cryotron gain curves using indium gates.

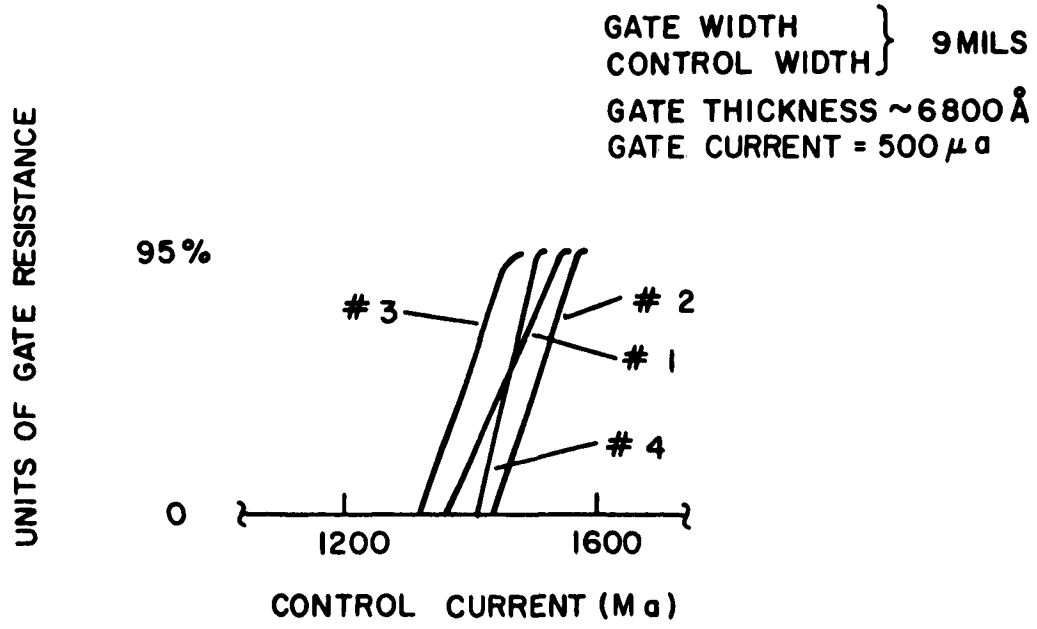


Figure 1b - Gate resistance as a function of control current for each cryotron in Fig. 1a.

The results of Table II show that the critical fields of indium were reproducible across the substrate. The spread in the critical currents and control currents is much larger. These results reflect the limitations of reproducibility that are believed to be caused by either shadowing in the control lines or an inability to control the line width in the mask. It should be possible, however, to eliminate these variations.

It is believed that a nonuniform magnetic field from the control causes the control current transition widths to be broader than the external field transitions. The fact that the control current transition widths, $\Delta I_c/I_c$, are larger than the magnetic field widths, $\Delta H/H$, makes the circuit operating tolerances more severe.

A reproducibility sample was made using tin as a gate material and the characteristics of the tin were reproducible across the substrate. The broad transitions, however, made the circuit operating conditions unsatisfactory.

The square in Fig. 1a outlines the operating area for the four cryotrons connected in a circuit and biased from a common source. The lower horizontal leg of the square is the current required to switch the maximum resistance into a cryotron and the left vertical leg represents the amount of gate current available as

a control current. The control current transitions for 95 percent full resistance of each cryotron are shown in Fig. 1b. It may be seen that the area does not allow any margin of safety for either an overdrive in control current beyond the point at which 95 percent of the full resistance is achieved, or variations in the bias setting. Although the reproducibility will allow for multiple operation, it would not be sufficient for reliable high-speed circuit operation.

CRITICAL CURRENTS OF TIN AND INDIUM FILMS

The dc critical currents were measured for a number of tin and indium films. The experimental values are compared with the predicted values calculated from a simple model for a thin film based on the London field and current distributions. The maximum current density at the surface of a gate film above a superconducting ground plane is

$$j_{max} = \frac{I_g}{W_g} \beta \coth \beta t, \qquad (1)$$

where j_{max} = maximum current density at the film surface (amp/cm^2),

I_g/W_g = critical gate current per unit gate width (amp/cm),

β = the reciprocal effective penetration depth (cm^{-1}),

t = film thickness (cm).

Eq. (1) has been derived previously[4] in essentially the same form based on the critical current density switching hypothesis. The critical current density switching hypothesis assumes switching when

$$j_{max} = \beta H_c^b, \tag{2}$$

where H_c^b = critical field of the bulk material. Hence, Eq. (1) becomes

$$\frac{I_g}{W_g} \frac{1}{H_c^b} = \tanh \beta t \ . \tag{3}$$

The symbols for Eq. (3) are the same as those for Eqs. (1) and (2). Eq. (3) states that the ratio of the critical current of a thin film to the critical current of a thick film should be tanh βt.

The experimental results for a number of tin and indium films are compared with Eq. (3) in Fig. 2 and 3. The critical current for these films is defined as the gate current that produces the smallest detectable voltage across the film network as the film becomes normal. The smallest detectable voltage is in the range of 0.1 to 0.5 microvolt and represents a resistance of 10^{-5} to 10^{-6} ohms.

The effective penetration depth values as inferred by W. B. Ittner[5] were used for the calculations of Eq. (3) and all the other necessary calculations. The experimental results were taken

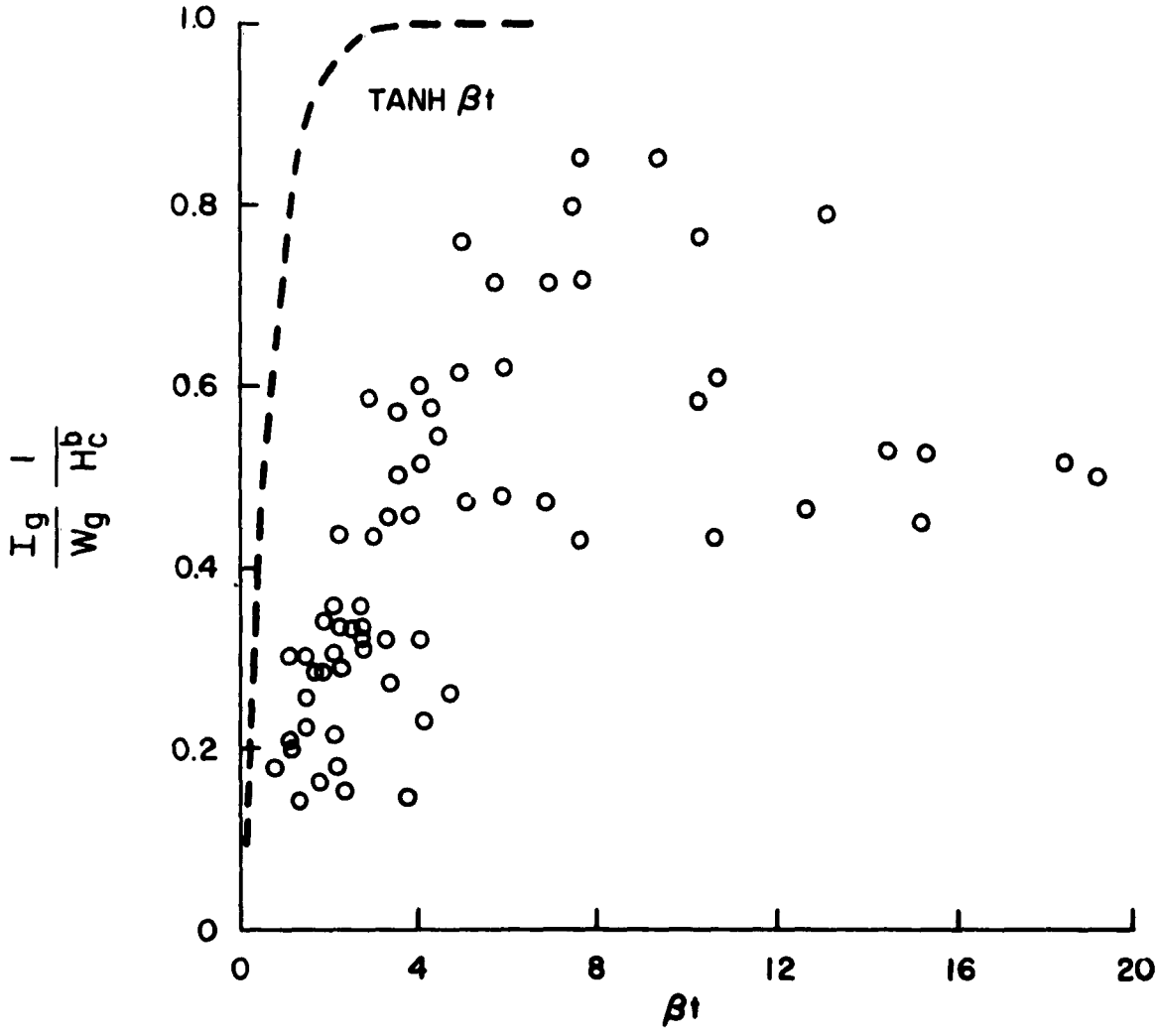


Figure 2 - Critical current of thin superconducting tin films normalized with respect to the critical current of a very thick film.

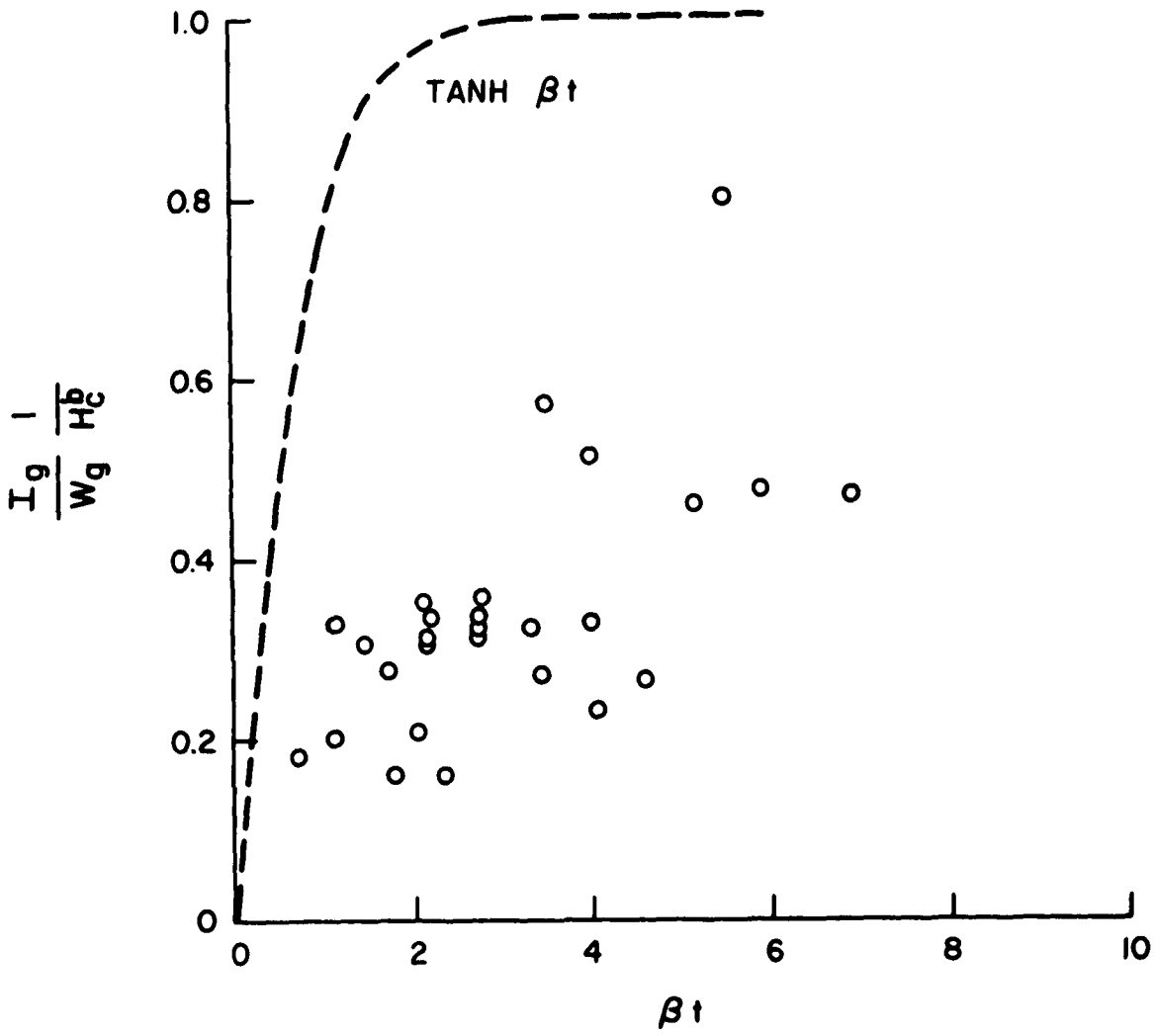


Figure 3 - Critical current of thin superconducting indium films normalized with respect to the critical current of a very thick film.

from a large number of films chosen at random from those produced in conventional evaporators.

The results shown in Fig. 2 for tin films seem to have a trend that is in qualitative agreement with those predicted by Eq. (3). The results for indium in Fig. 3 also have the trend predicted by Eq. (3), but the values are smaller than those of tin.

It is not expected that the experimental results would be those predicted by Eq. (3) because of the number of broad assumptions in the critical current density switching hypothesis, such as the current distribution in a film and the manner in which current switching occurs. Also some physical properties of the films are not considered, such as grain size and irregularities in the film dimensions.

GAIN CHARACTERISTICS OF CROSSED-FILM CRYOTRONS

The gain characteristics were taken on a number of multiple-control unity-crossing crossed-film cryotrons. Cryotrons with both tin and indium gates were evaporated in conventional systems.

The gain characteristics for these cryotrons were all similar to those reported in the eighth quarterly Lightning report, pages 18 and 20. There is incremental gain over a small range of gate currents at lower temperatures. The range of incremental gain, however, is so limited that these cryotrons would be impossible to operate if interconnected. Though the gate material characteristics of indium were good, the gain characteristics were still unsatisfactory.

The static gain of a number of crossed-film cryotrons is plotted in Fig. 4 as a function of βt. The static gain is the ratio of the intercepts of the crossed-film cryotron gain curve and the current axes. The gain has been discussed previously[4] and will be repeated in part below. The gain of a crossed-film cryotron is

$$G = \frac{I_g}{I_c}, \tag{4}$$

where I_g is the critical gate current with no control current, and I_c is the control current to switch the gate in the absence of gate current.

The gain may be expressed in analytical form with the aid of the London theory. The critical current density switching hypothesis is used to establish the dependence of critical gate current on normalized gate thickness (βt). London's free-energy approach is used

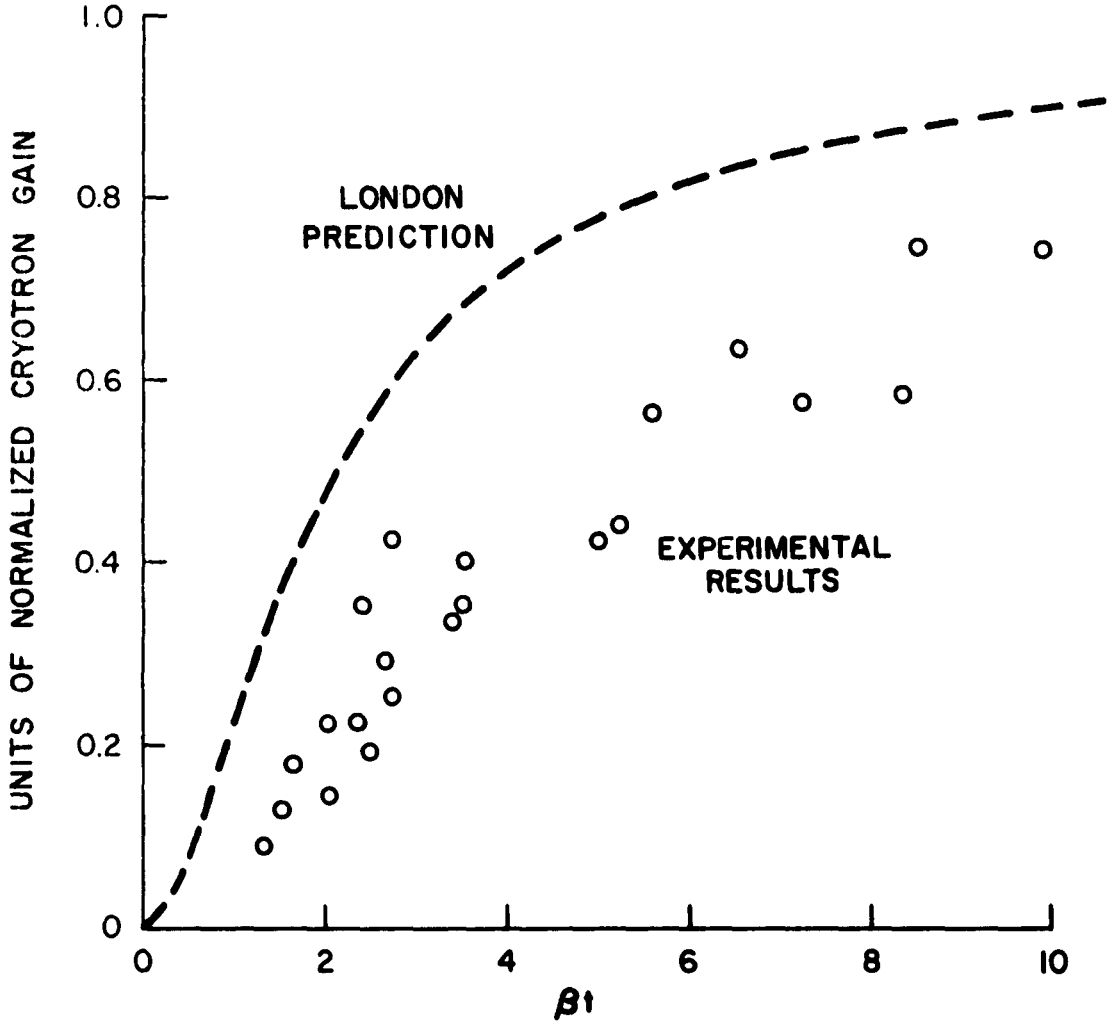


Figure 4 - The gain of crossed-film cryotrons. The dashed curve, derived from the London theory and the critical current density hypothesis, is compared with the experimental results as a function of βt.

to relate the critical control current to the critical field of bulk and βt. With these assumptions, the gain expression is

$$G = \frac{W_g}{W_c} \frac{H_c^b}{H_c^f} \tanh \beta t, \qquad (5)$$

where G = gain

W_g/W_c = ratio of the width of the gate to the width of the control

H_c^b/H_c^f = ratio of the bulk critical field to the film critical field

$\beta = 1/\lambda_\epsilon$ the reciprocal effective penetration depth

t = film thickness.

The London free-energy relation[6] for H_c^b/H_c^f is used in Eq. (5) to produce the normalized gain expression of Eq. (7). Eq. (6), the ratio H_c^f/H_f^b, has been developed from the London relation neglecting surface energy considerations. The development of Eq. (6) has been previously[7] shown to be

$$H_c^f/H_c^b = \frac{1}{\sqrt{1 - \frac{2}{\beta t} \tanh \frac{\beta t}{2}}} \qquad (6)$$

Inserting Eq. (6) into Eq. (5), the normalized gain becomes

$$G = \tanh \beta t \sqrt{1 - \frac{2}{\beta t} \tanh \frac{\beta t}{2}}. \qquad (7)$$

The expected results calculated from Eq. (7) are plotted in Fig. 4. The experimental results from a number of tin crossed-film cryotrons are compared with the expected results in Fig. 4. The experimental results are in qualitative agreement with the expected results but are generally 30 to 40 percent lower than the predicted values. The discrepancy is not extremely large when considering the limited accuracy in the experiments and the broad assumptions used in deriving the analytical expression.

It is believed that the results in Fig. 4 may give some useful information for the design of cryotron circuits using crossed-film cryotrons.

CONCLUSIONS

Indium produced in the conventional device systems was satisfactory for device work, while tin from the same systems suffered from the "edge effect." Indium characteristics are reproducible from independent evaporations, but the variations in control line width were too large to make cryotrons with characteristics acceptable for high-speed circuit operation. The characteristics of tin from a single evaporation are reproducible, but the broad magnetic transitions of the tin when deposited on unheated substrates

reduce the margins of safety for satisfactory circuit operation. The characteristic gain curve of a crossed-film cryotron is still neither predictable nor reproducible. Furthermore, it does not provide incremental gain greater than unity over an adequate current range, with equal gate and control widths. The gain values of a crossed-film cryotron seem in qualitative agreement with the gain values predicted by use of a simple theory involving the London field and current distributions.

REFERENCES

1. A. Toxen, "Size Effects in Thin Superconducting Indium Films," to be published, Phys. Rev., July 15, 1961.

2. H. Caswell, IBM Research, private communication.

3. Project Lightning, Eighth Quarterly Progress Report, pp. 18-20.

4. Project Lightning, Second Quarterly Progress Report, pp. 28-29.

5. W. B. Ittner, III, Phys. Rev., Vol. 119, p. 1591, September, 1960.

6. F. London, "Superfluids," John Wiley & Sons, Inc., New York, 1950, Vol. 1, p. 131.

7. Project Lightning, Fourth Quarterly Progress Report, pp. 127-128.

APPENDIX II

Time Average Thermal Properties of a Computer Utilizing Thin-Film Superconducting Elements

H. Sobol

ABSTRACT

It is necessary to understand the static thermal properties of cryotron gates before one can predict the limitations associated with dissipative heating of a complex cryotronic computer. An experimental program has been conducted to determine the thermal properties of isolated and of thermally coupled gates. All experiments reported in this paper were performed on tin gates evaporated onto glass substrates.

The total thermal conductance, K, of a gate is defined, and experimental values of K are presented as a function of power and gate geometry. An analysis of the heat flow is given, based on temperature-dependent coefficient of heat transfer. Theoretical values of K and temperature distribution are derived. The theory qualitatively predicts the temperature variation. Finally, the results are extrapolated to estimate the number of cryotrons that can be used safely in a thermally coupled system.

INTRODUCTION

Thin-film superconducting components offer a means of realizing a computer capable of high speed and low dissipation operation with large packing densities. While the temperature rise produced by the expected low dissipation may be insignificant to devices operating at room temperature, this same rise may have a profound effect on the operation of a superconducting device, operating at liquid helium temperatures. The penetration depth of a magnetic field into a thin superconducting film is a function of temperature and therefore inductance (and hence switching time) and cryotron gain will vary with temperature. In order to insure satisfactory operation, the ambient or average temperature rise of a plane of circuitry must be kept below a specified value.

The factors determining the temperature rise are the average power dissipation in the individual cryotron circuits, the thermal coupling between cryotrons, and the heat transfer between the plane of circuitry and the helium bath.

In an earlier work, Ittner,[1] using a relatively simple model, calculated the limit on the number of cryotrons per substrate imposed by dissipation during the switching operations. The model assumed an isothermal plane of circuitry and a dissipation based on

very high frequency operation. At the time of his writing, the only data available on the heat transfer process from a substrate to a liquid helium bath were results of preliminary experiments performed at IBM and published data on the heat transfer between metallic surfaces and Helium I.

It is the purpose of this paper to report on an investigation of the static thermal properties of evaporated thin-film cryotron gates* on glass substrates. The heat transfer between the gates and a Helium I bath is studied. Using the results of the study, it is possible to estimate cryotron packing densities in terms of a limit on the ambient temperature rise, electrical properties, frequency and duty factor, and geometry. An estimate is also given for the relative gain in packing density resulting from the use of high thermally conducting substrates.

The dynamic thermal behavior of cryotron circuits operating at high repetition rates is the subject of another study. The results[2] have been presented in the literature. It was shown that when the electrical switching speed is faster than the thermal response, the gate temperature increases by an amount almost proportional to the average power dissipated, and is practically

---

* Gates are usually the only dissipative elements in cryotron circuits.

independent of time. This justifies using the static thermal properties to determine the limitations of a computer operating at high frequencies.

In an earlier report on the thermal properties of cryotrons,[3] a coefficient of heat transfer to a Helium I bath was determined by substituting data obtained from measurements of the temperature distribution across a glass substrate into a linear one-dimensional model, assuming a constant heat-transfer coefficient. The exponential character of the temperature distribution was predicted by the linear theory, but the heat-transfer coefficient was lower than published values[4] for the transfer of heat from a metallic surface. Using a model with a temperature-dependent heat-transfer coefficient, h, it is shown that h may be as large as the published values in the vicinity of the gate but falls to the magnitude of values previously found at positions distant from the gate.

An attractive method for reducing the thermal problem would be to operate at temperatures below the λ point of helium ($2.2^{\circ}K$), and take advantage of the very large heat-transfer coefficient obtainable with superfluid helium. Unfortunately, the refrigeration problem becomes very difficult below the λ point. Also, to keep the current levels from getting prohibitively high, it would be

necessary to use new gate materials with low critical temperatures. The immediate problem therefore is to study the thermal limitation associated with a computer operating at Helium I temperatures.

TOTAL THERMAL CONDUCTANCE OF AN ISOLATED GATE

Let us now consider the static thermal properties of a cryotron gate evaporated onto a substrate and immersed in a liquid helium bath. The total thermal conductance, K, of the gate relates the ambient temperature rise, ΔT_{av}, to the average power, P_{av}, dissipated in the gate. This conductance is defined in Eq. (1):

$$K \left(\frac{\text{watts}}{^{o}\text{Kelvin}} \right) = \frac{P_{av}}{\Delta T_{av}} . \qquad (1)$$

The conductance, K, depends on the bath temperature, the temperature rise of the gate and its surroundings, the conductivity and geometry of the substrate and of all films between the gate and the bath, the coefficient of heat transfer at each interface surface, and the gate geometry. Most of these thermal parameters are virtually unknown and must be determined experimentally.

A photograph of the sample used in the measurement of K is shown in Fig. 1, and a cross-sectional view of the film layers is shown in Fig. 2. The sample has a lead (Pb) film ground plane covering most of the glass substrate. Twelve tin film gates are evaporated over the ground plane. Silicon monoxide (SiO) is used as the insulating material and lead (Pb) connecting lines are used. Five gates are 0.0106 inch wide and vary in length from 1/16 to 5/16 inch. Four gates are 3/16 inch long and vary in width from 0.0045 to 0.0315 inch. The last three gates vary in length and width but have approximately the same area.

The total conductance of each gate was determined as a function of the power dissipated in the gate. The thermal hysteresis method of measuring K was used. This method has been described previously.[5] Briefly, it consists of running gate thermal hysteresis loops at various bath temperatures. K is then the ratio of the power dissipated to the temperature difference between the superconducting-to-normal and normal-to-superconducting phase transitions at each gate-current level.

The results of the experimental measurements of K are shown in Fig. 3, 4, and 5. These curves are the mean values for several runs on three samples. A typical scattering of points is

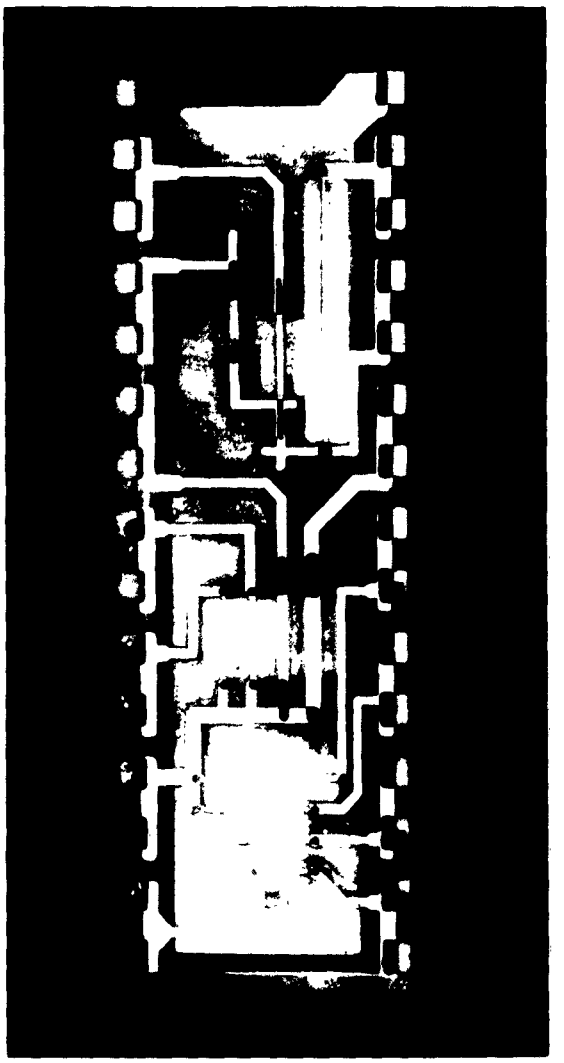
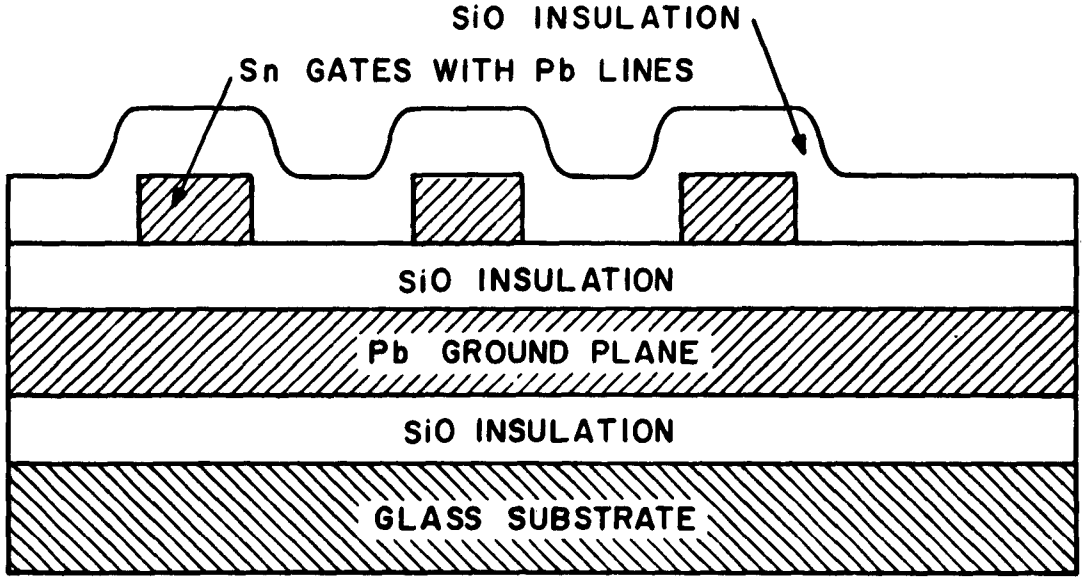


Figure 1 - Photograph of sample used for measuring K (glass substrate, 2 x 5/8 x 0.04 in.).



SUBSTRATE DIMENSIONS 5/8 × 2 × 0.040 IN.

FILM THICKNESS (Å)

SiO	5000
Pb	10 000
Sn	6000

Figure 2 - End view of sample for K measurements (not to scale).

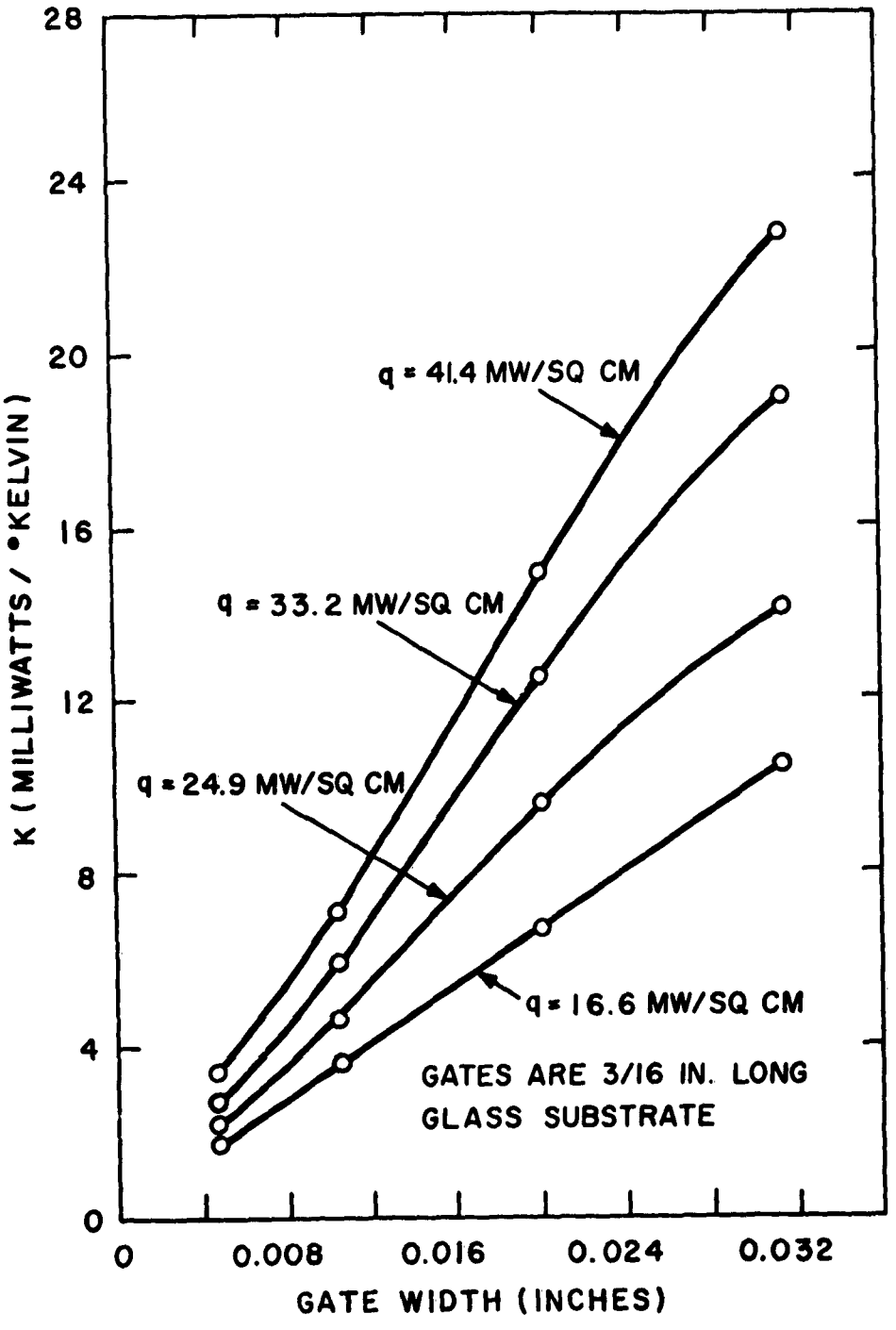


Figure 3 - Mean values of K as a function of gate width for fixed length with power per unit area as parameter.

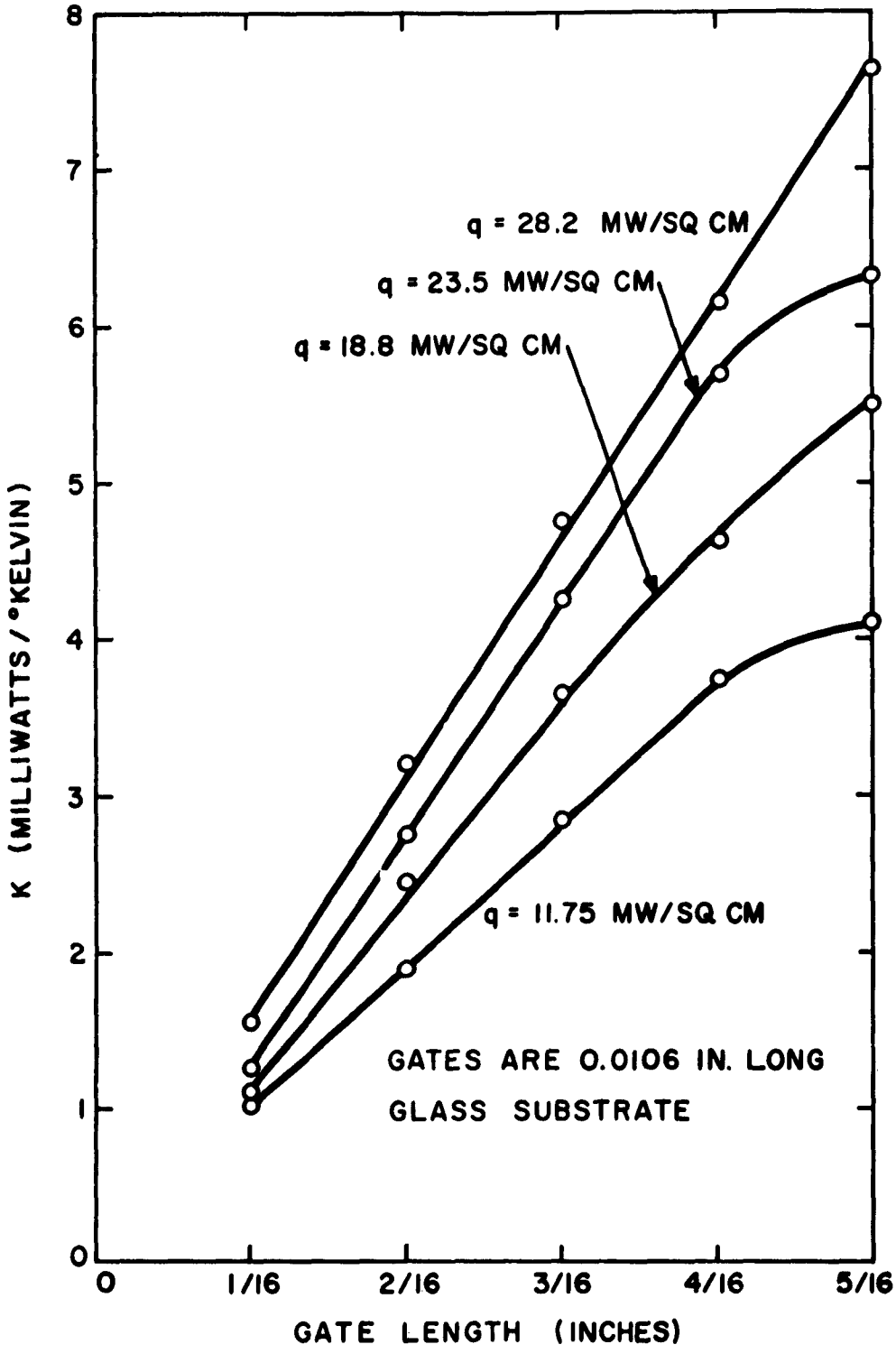


Figure 4 - Mean values of K as a function of gate length for fixed width with power per unit area as parameter.

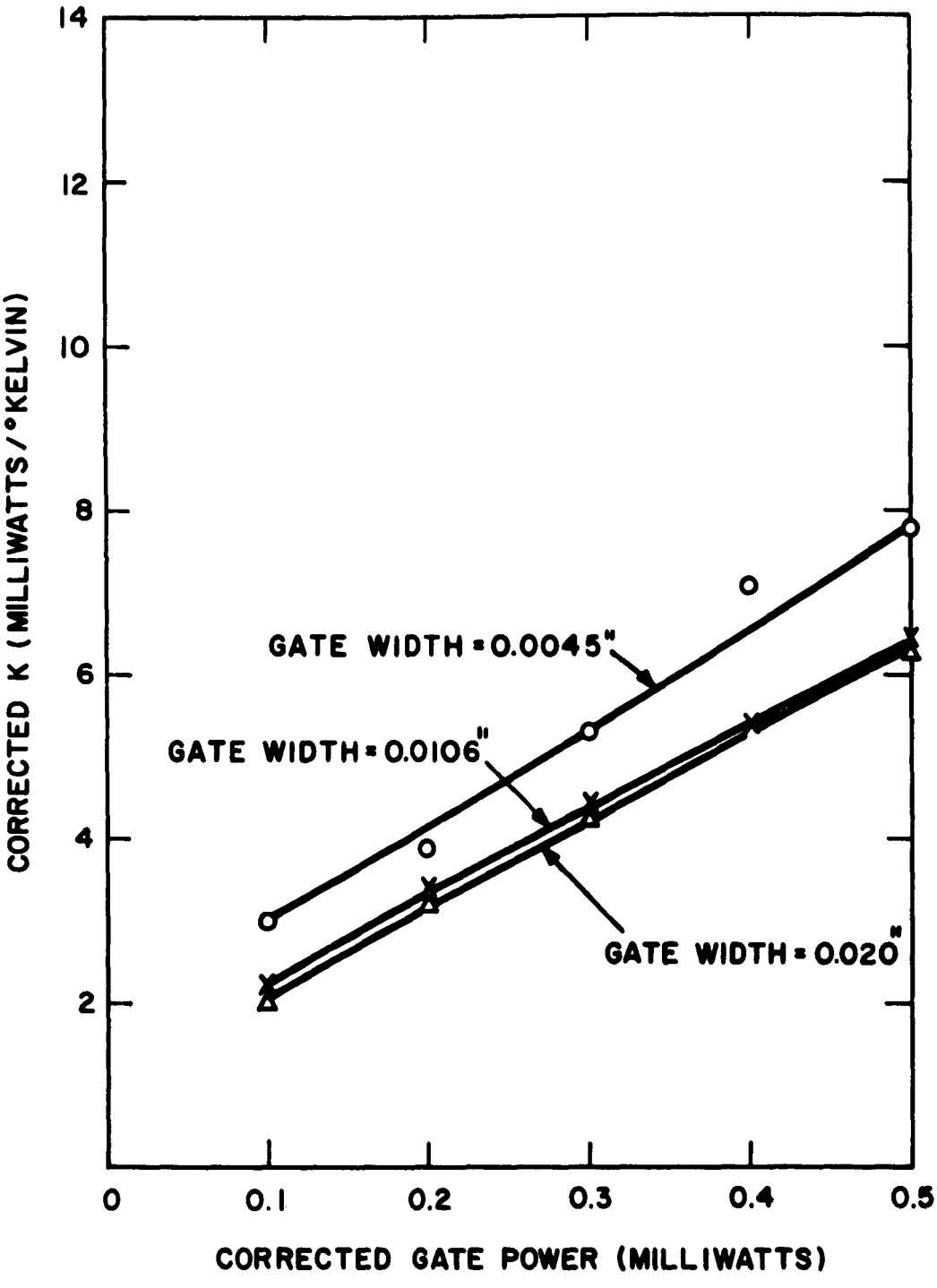


Figure 5 - Mean values of K as a function of gate power.

shown in Fig. 6. The spread of points is about 25 percent in the worst case. The reasons for the spread are general fabrication differences between samples, the inability to measure temperature differences accurately by the hysteresis method,* gate width variations, and irreproducibility of critical current measurements.

Figures 3 and 4 indicate that for a specific power per unit area dissipated by the gate, K varies almost linearly with length and width for fixed width and length, respectively. The area was not constant, however, for the "constant area" gates. The error was caused by oversized masks used during the evaporation process. The calculated K and measured power were "corrected" by multiplying each by the ratio of desired area to the actual area. While the curves of Fig. 5 are fairly close to each other, it is to be noted that the narrow gates have a somewhat larger K. This will be discussed in more detail later. From Fig. 3 and 4, it can be concluded that K varies linearly with gate area (for fixed power densities), and the following equation is postulated to describe the total thermal conductance:

$$K = \frac{P_{av}}{\Delta T} = fA_g h \ . \qquad (2)$$

---

* It is thought that this method can determine temperatures to within about 10 millidegrees.

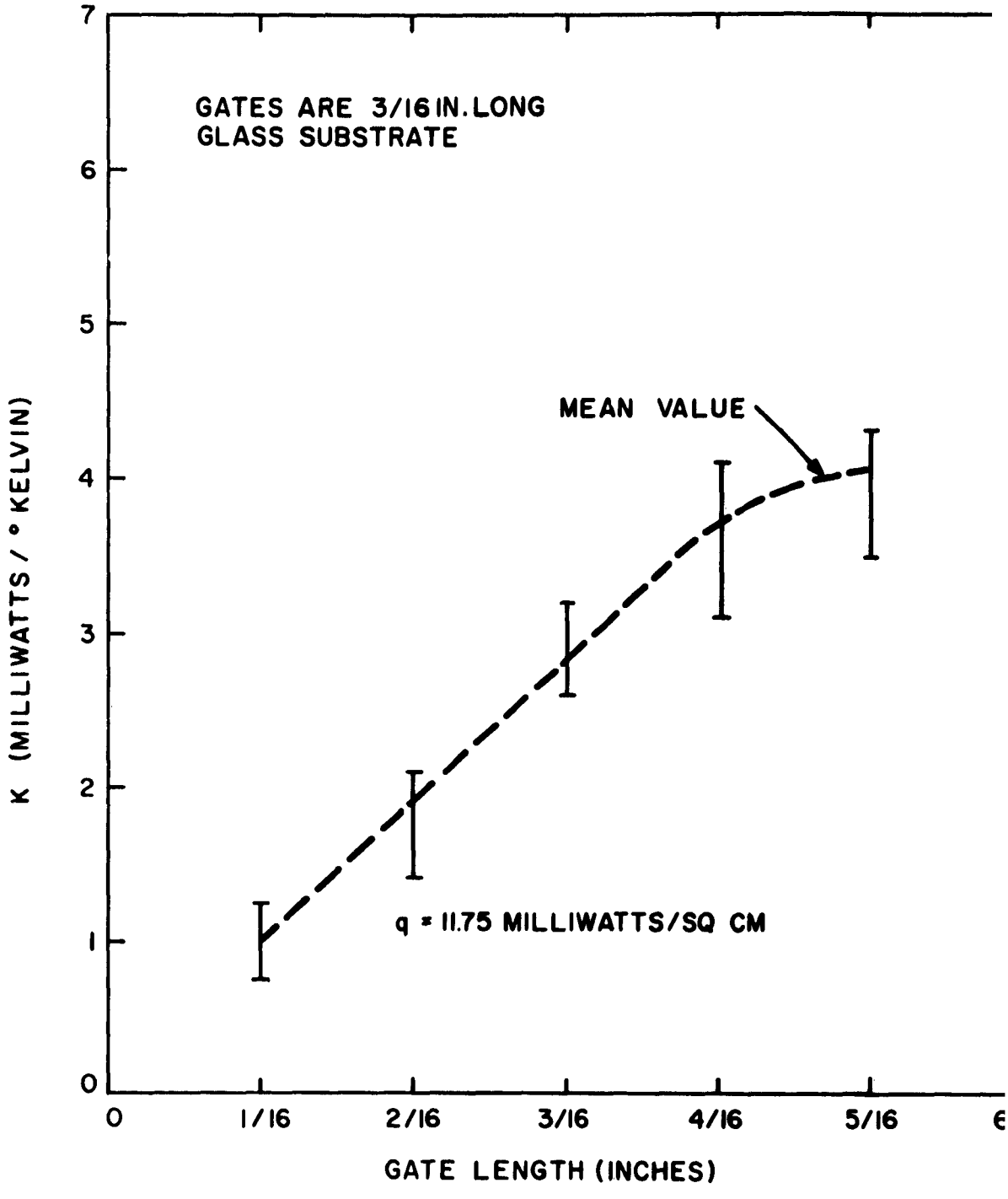


Figure 6 - K as a function of gate length.

A_g is the geometrical area of the gate, and the product fA_g is the effective area over which heat is transferred to the bath, and h is the coefficient of heat transfer to the helium at the gate temperature. In most experiments for determining h, a simple geometry is used and the heat flux enters and leaves through a known area. In the present experiment, the situation is very complex since the heat flux is generated in the gate but returns to the bath over portions of the entire substrate. Because of this complication, the "averaging" factor f must be included.

The factor f can also be associated with h and the product fh expressed as

$$fh = \frac{P/A_g}{\Delta T} = \frac{q}{\Delta T} \qquad (3)$$

The power density q was determined by gate resistance and current measurements. The measured temperature rise ΔT was averaged over all 0.0106-inch gates on three samples, at specific values of q. Figure 7 shows a plot of the average q-ΔT plane. The average values of fh can then be determined as a function of the temperature difference by using Fig. 7 and Eq. (3). The result is shown in Fig. 8. The bath temperature was varied between 3.55 and 3.76°K for the points plotted in Fig. 7 and 8.

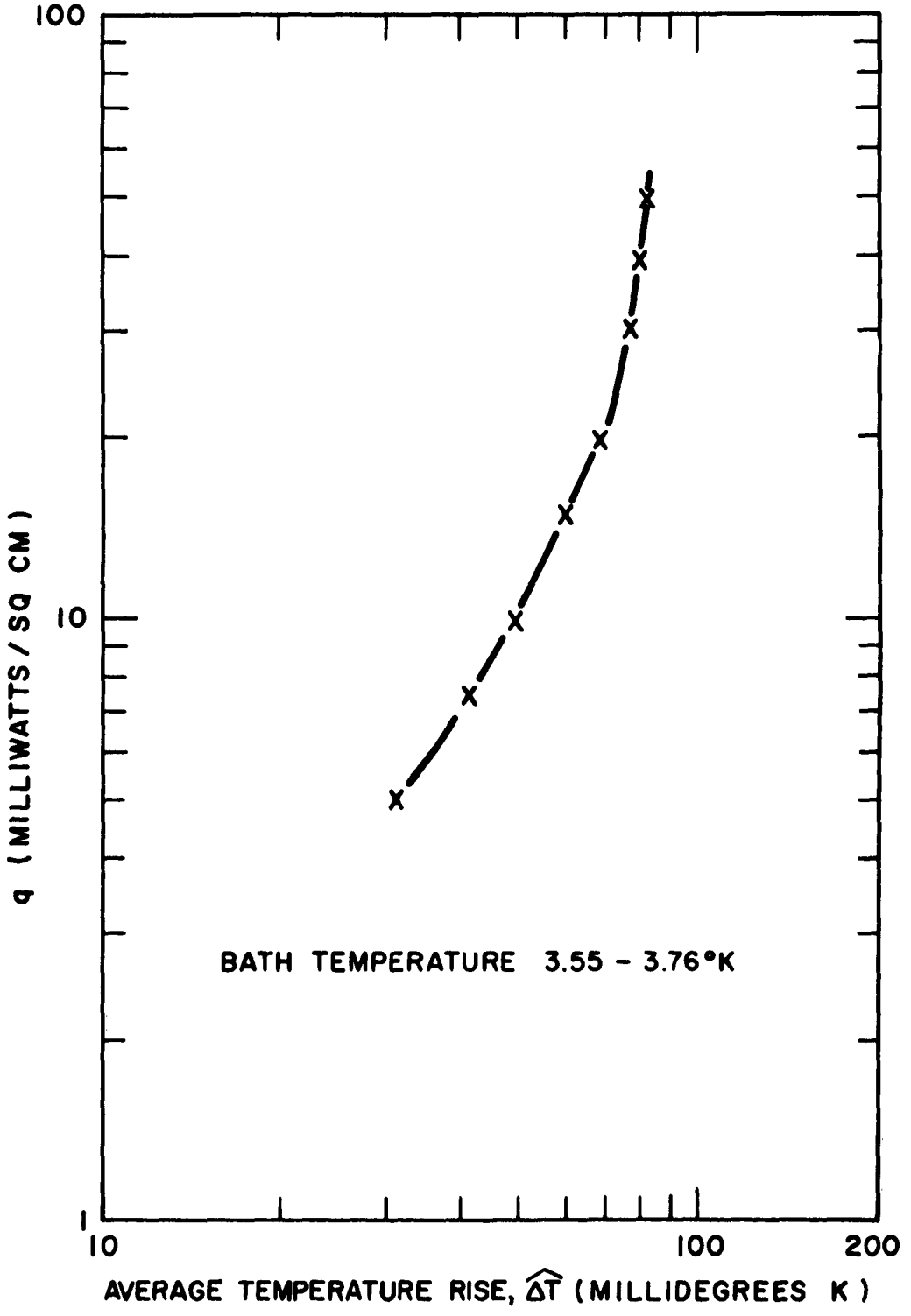


Figure 7 - Power per unit gate area as a function of temperature rise.

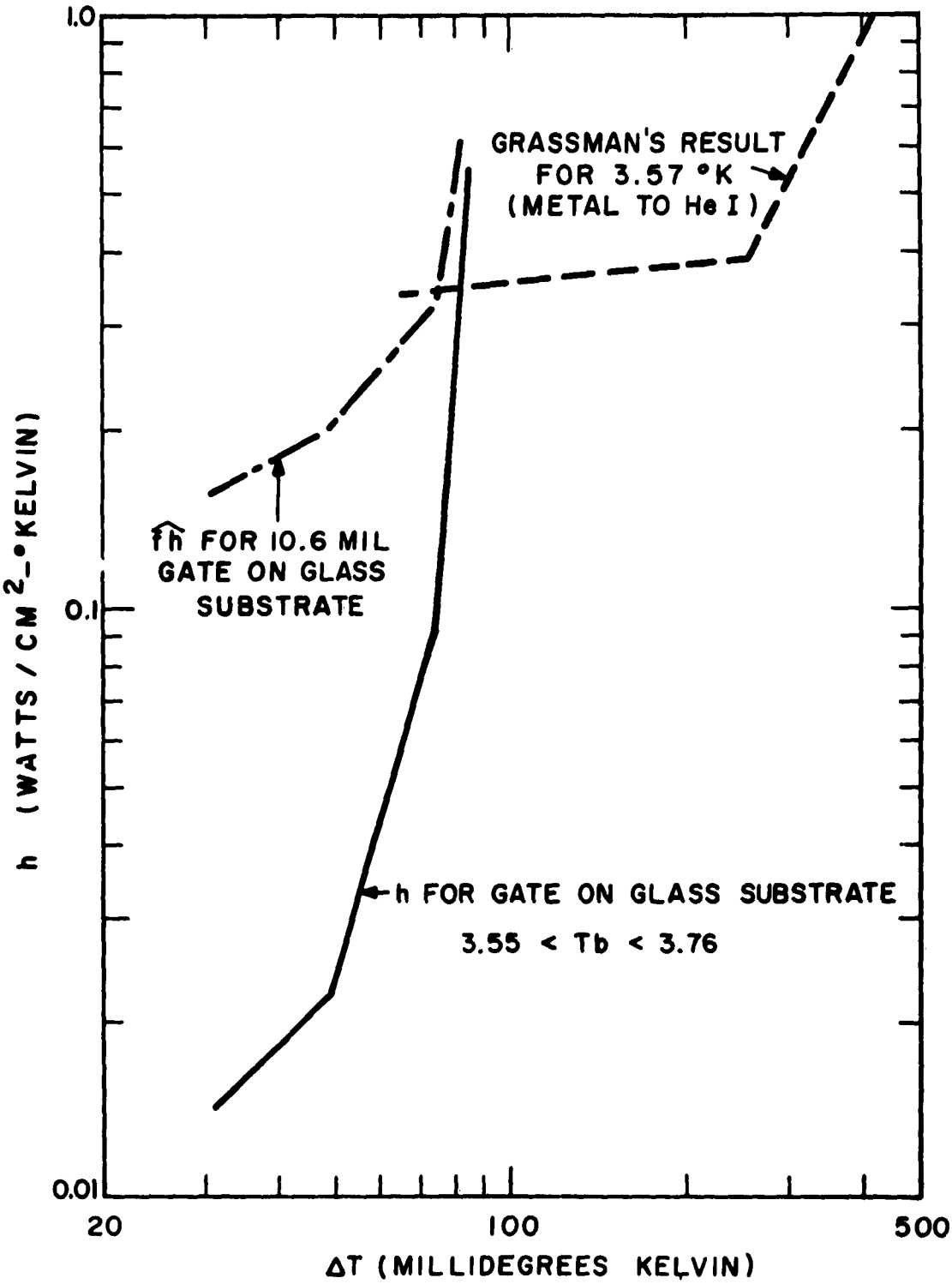


Figure 8 - Coefficient of heat transfer to He I bath as a function of temperature rise.

The product fh will now be separated so that we may determine the coefficient of heat transfer h. In order to do this, it is necessary to calculate the theoretical temperature variation across the substrate.

TEMPERATURE DISTRIBUTION

A one-dimensional analysis will be used to determine the space variation of temperature. The model used is shown in Fig. 9. The temperature drop across the insulating films is neglected.* The ground plane and substrate are replaced by an equivalent substrate having the same thickness as the glass but twice the conductivity.**

The static temperature distribution along the substrate is governed by

$$\frac{d^2 \Delta T(x)}{dx^2} = \frac{2h\,(\Delta T)\quad \Delta T(x)}{k\delta} , \qquad (4)$$

---

* This assumption holds for glass substrates but may be invalid with sapphire or aluminum.

** The product of the conductivity and the thickness is approximately the same for the superconducting lead ground plane and the glass substrate for liquid helium temperatures.

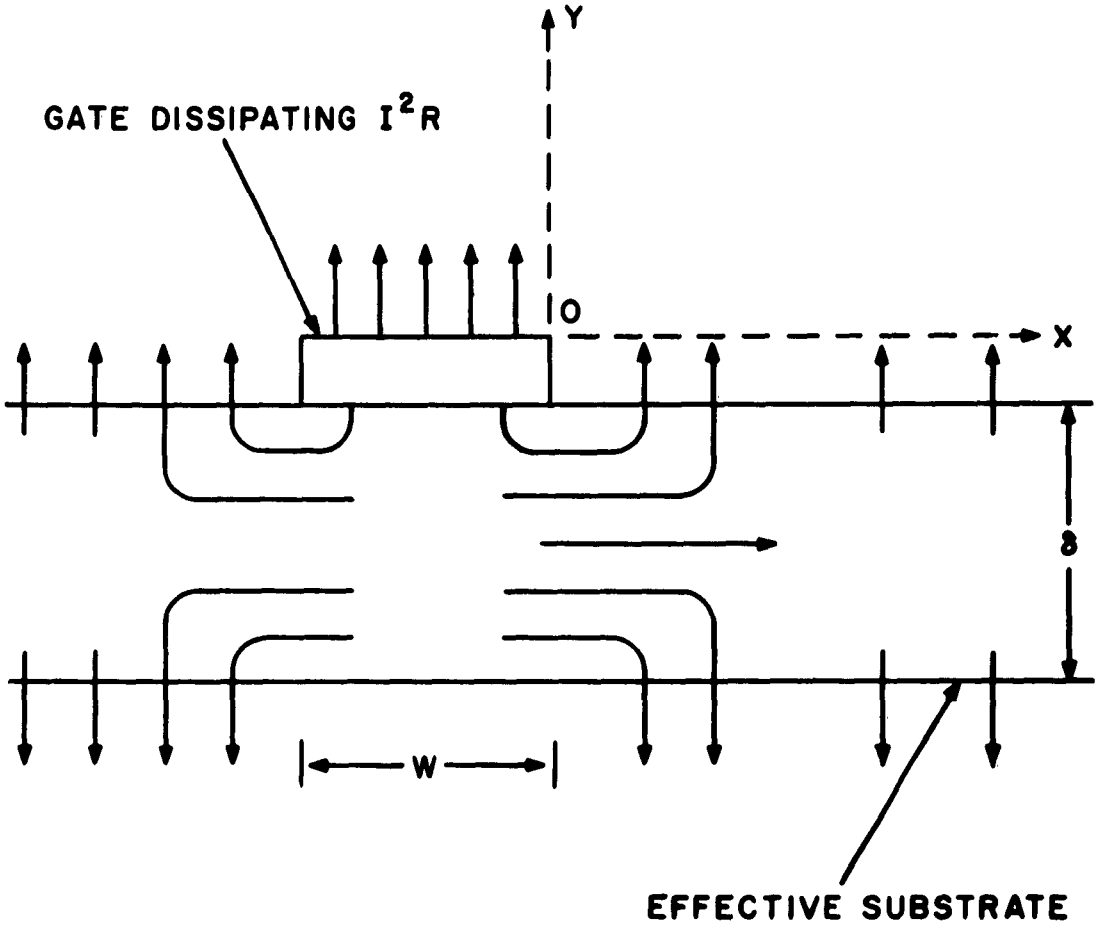


Figure 9 - Model for analysis of heat flow.

where x is measured as the distance from the edge of the gate, k is the thermal conductivity of the equivalent substrate, and δ is the thickness of the glass. Let us assume that the transfer coefficient can be described as

$$h = h_0 \left(\frac{\Delta T}{\Delta T(0)} \right)^m , \qquad (5)$$

where h_0 corresponds to the transfer coefficient at the temperature difference $\Delta T(0)$ and m is an exponent to be determined. The motivation for this particular choice of temperature dependence was based on the variation of fh with temperature indicated in Fig. 8. Since three distinct transfer processes are indicated in Fig. 8, it is assumed that the transfer coefficient will have the three ranges:

$$h_1 = h_0 \left(\frac{\Delta T}{\Delta T(0)} \right)^{m_1} \quad \text{for } \Delta T > 73 \text{ millidegrees,}$$

$$h_2 = h_0' \left(\frac{\Delta T}{\Delta T(0)} \right)^{m_2} \quad \text{for } 49 < \Delta T < 73 \text{ millidegrees,}$$

$$h_3 = h_0'' \left(\frac{\Delta T}{\Delta T(0)} \right)^{m_3} \quad \text{for } \Delta T < 49 \text{ millidegrees.}$$

The change from process h_1 to h_2 will occur at distance x_{b_1} and the change from h_2 to h_3 will occur at x_{b_2}.

A piece-wise solution to Eq. (4), under the conditions that the temperature at the gate is $\Delta T(0)$, $\Delta T(\infty) \rightarrow 0$, and that the temperature is continuous at x_{b_1} and x_{b_2}, is given as:

for $0 < x < x_{b_1}$,

$$\frac{\Delta T}{\Delta T(0)} = \left[1 + \frac{\sqrt{\frac{2h_0}{k\delta}}\, x}{\sqrt{\frac{2}{m_1}\left(\frac{2}{m_1} + 1\right)}} \right]^{\frac{-2}{m_1}} , \quad (6)$$

for $x_{b_2} < x < x_{b_1}$,

$$\frac{\Delta T}{\Delta T(0)} = \frac{0.073}{\Delta T(0)} \left[1 + \frac{\sqrt{\frac{2h_0'}{k\delta}}\, (x - x_{b_1})}{\sqrt{\frac{2}{m_2}\left(\frac{2}{m_2} + 1\right)}} \right]^{\frac{-2}{m_2}} , \quad (7)$$

and for $x > x_{b_2}$,

$$\frac{\Delta T}{\Delta T(0)} = \frac{0.049}{\Delta T(0)} \left[1 + \frac{\sqrt{\frac{2h_0''}{k\delta}}\, (x - x_{b_2})}{\sqrt{\frac{2}{m_3}\left(\frac{2}{m_3} + 1\right)}} \right]^{\frac{-2}{m_3}} , \quad (8)$$

where h_0 is the transfer coefficient at $\Delta T(0)$,

h_0' is the transfer coefficient at $0.073^{\circ}K$, and

h_0'' is the transfer coefficient at $0.049^{\circ}K$.

The positions x_{b_1} and x_{b_2} are given by

$$x_{b_1} = \sqrt{\frac{k\delta}{2h_0}} \sqrt{\frac{2}{m_1}\left(\frac{2}{m_1} + 1\right)} \left\{ \left[\frac{\Delta T(0)}{0.073}\right]^{\frac{m_1}{2}} - 1 \right\} \quad (9)$$

and

$$x_{b_2} = \sqrt{\frac{k\delta}{2h_0'}} \sqrt{\frac{2}{m_1}\left(\frac{2}{m_2} + 1\right)} \left\{ \left[\frac{0.073}{0.049}\right]^{\frac{m_2}{2}} - 1 \right\} . \quad (10)$$

Eqs. (6) through (10), together with the fact that the gate itself is isothermal at temperature difference $\Delta T(0)$, describe the space variation of temperature. We will now determine the total thermal conductance.

THEORETICAL K

The heat-balance equation under static conditions is given by

$$P_{av} \simeq A_g h_0 \Delta T(0) + \frac{2\ell\, h_0}{\Delta T(0)^{m_1}} \int_0^{x_{b_1}} \Delta T(x)\Big]^{m_1+1} dx$$

$$+ \frac{2h_0' \ell}{\Delta T(0)^{m_2}} \int_{x_{b_1}}^{x_{b_2}} \Big[\Delta T(x)\Big]^{m_2+1} dx$$

$$+ \frac{4h_0'' \ell}{\Delta T(0)^{m_3}} \int_{x_{b_2}}^{\infty} \Big[\Delta T(x)\Big]^{m_3+1} dx$$

$$+ \ell h_0'' \left[w + 2x_{b_2}\right] \Delta T(x_{b_2}) \quad . \tag{11}$$

In writing this heat-balance equation it has been assumed that a gate ℓ units long and w units wide is on a substrate that extends infinitely in the plus and minus x directions. The first term on the right-hand side of Eq. (11) describes the heat transferred from the gate directly to the bath. The second and third terms

describe the heat transferred from the top surface of the substrate in the region $0 < x < x_{b_2}$. The fourth term describes the heat transferred from the upper and lower surfaces of the substrate for $x > x_{b_2}$. The last term describes the heat transfer from the lower surface of the substrate in the range covered by the gate and extending to $\pm x_{b_2}$. It is assumed that the substrate area below the gate is not isothermal in the y direction and that the temperature along the lower surface of the substrate in this region is $\Delta T(x_{b_3}) \approx 49$ millidegrees.

Eq. (11) can be put into the form

$$\frac{P}{A_g h_0 \Delta T(0)} = f \quad . \tag{12}$$

Using Eq. (6) through (11), and the measured values of fh given in Fig. 8, we can determine, after a laborious calculation, the exponents m_1, m_2, and m_3 and thus the values of h and f. The results are shown in Fig. 8 and 10, respectively.

In Fig. 11 and 12, theoretical temperature distributions are shown to agree fairly well with the experimental results presented in an earlier report.[6] It is to be noted that the first experimental point in all cases is always at a higher temperature than the theoretical value. This is a consequence of the one-dimensional

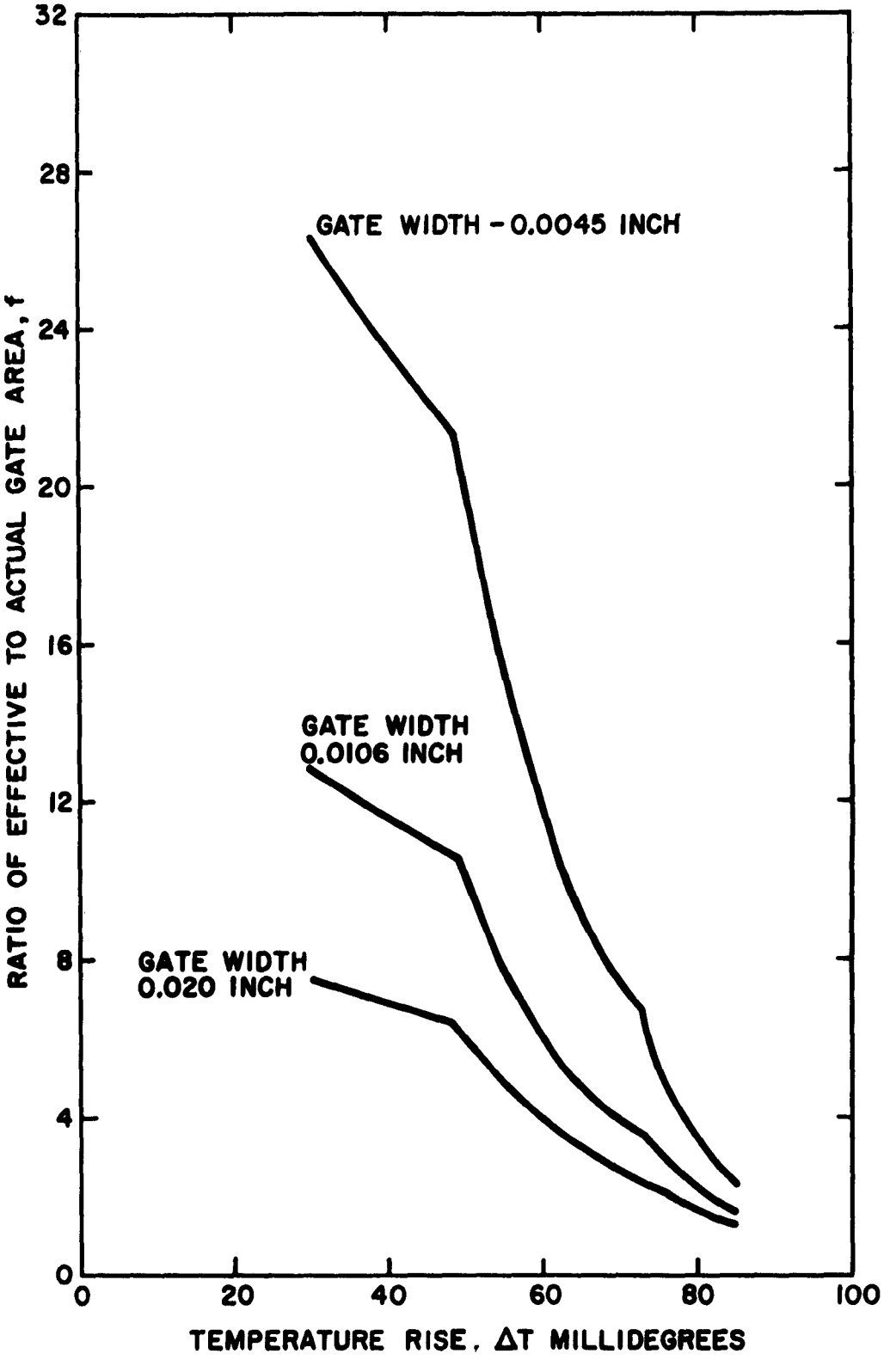


Figure 10 - Ratio of effective to actual gate area as a function of temperature rise (glass substrate).

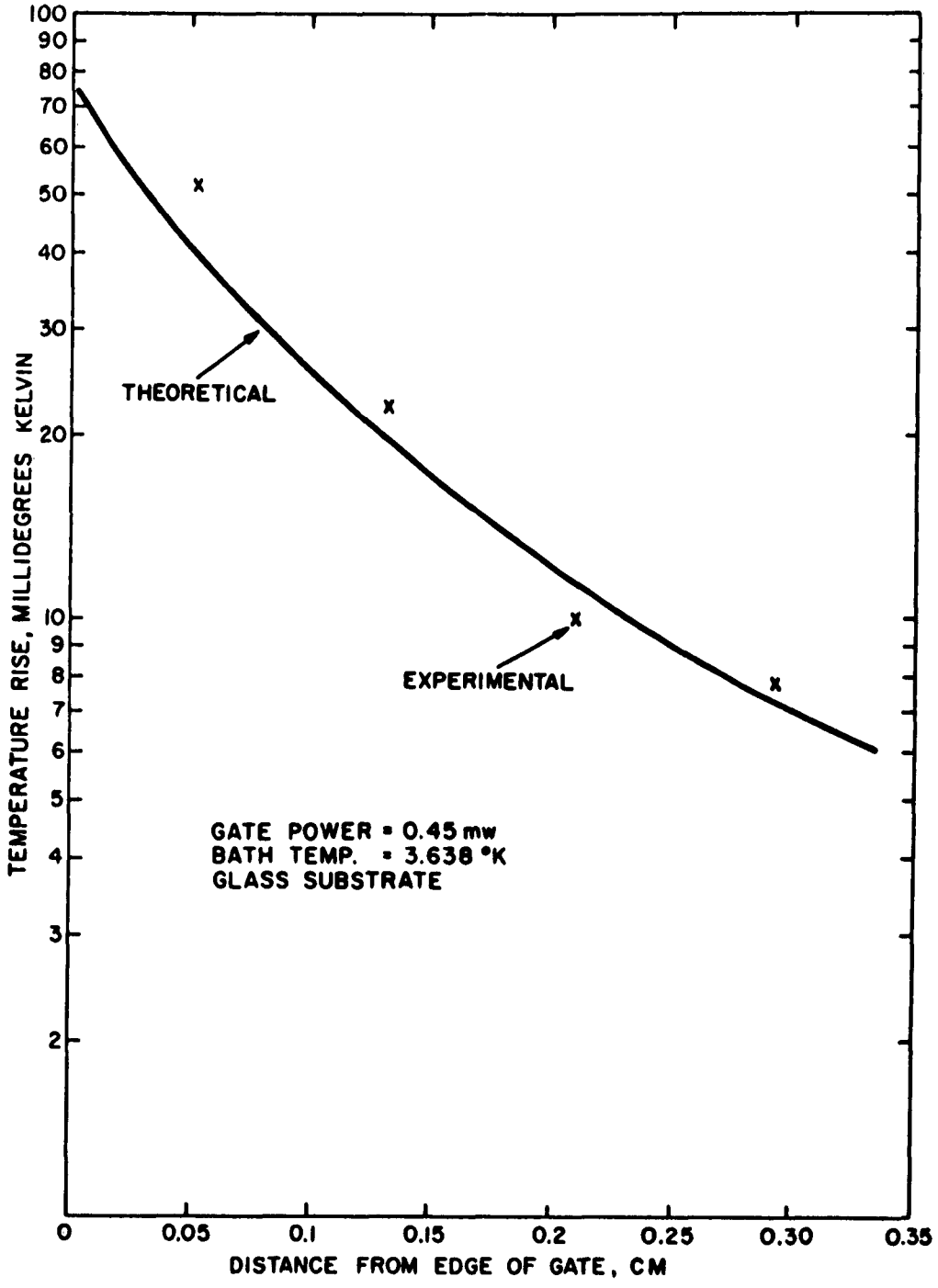


Figure 11 - Comparison of experimental and theoretical temperature distributions on sample 1223.

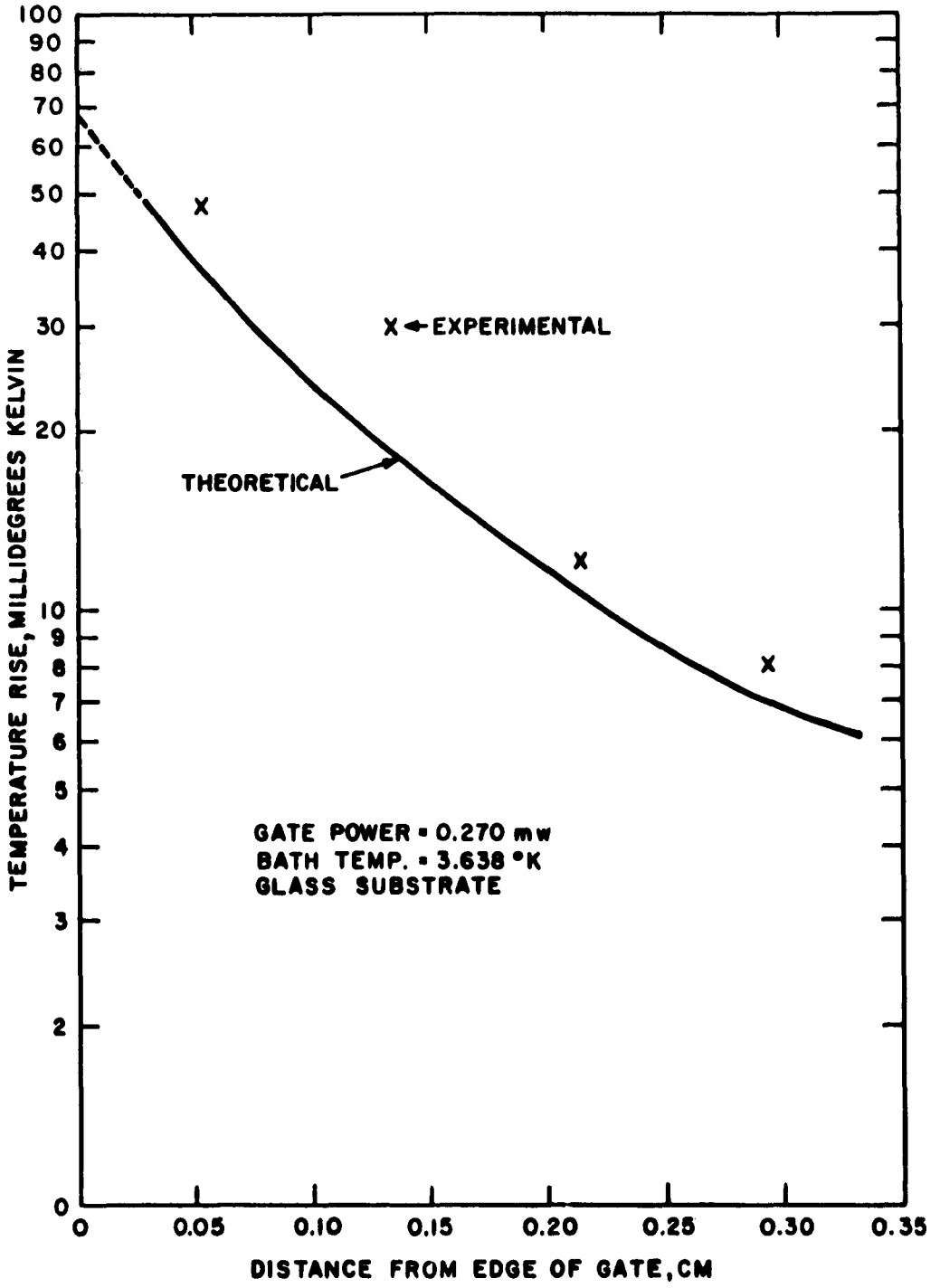


Figure 12 - Comparison of experimental and theoretical temperature distributions on sample 1105.

approximation. Some recent work[7] on a two-dimensional model has shown that the temperature along the top surface of the substrate, for a distance equal to the substrate thickness, is higher than the one-dimensional model predicts. At greater distances the substrate becomes isothermal in the y direction and a one-dimensional model predicts the same variation as the two-dimensional model.

The theoretical temperature distribution can be fitted to a first-order approximation with a straight line in the semi-log plots shown in Fig. 11 and 12. The slope of this line is given approximately as

$$a = \sqrt{\frac{2\,\frac{h_0''}{3}}{k\,\delta}} \quad . \tag{13}$$

Since h_0'' is equal to 22 mw/^{o}K cm^2, the slope indicated would correspond to an apparent transfer coefficient of about 7 mw/^{o}K cm^2. For the results indicated in Fig. 11, the transfer coefficient at the gate itself is 0.240 w/^{o}K cm^2. However, an estimate from the slope of the distribution curve would indicate an apparent transfer coefficient that is about 1/32 of the value at the gate.

This brings to focus the earlier dilemma of the very low coefficient of heat transfer determined from the original thermal

experiments. The difficulty is that a linear analysis was formerly used and this method cannot describe adequately this nonlinear problem.

COEFFICIENT OF HEAT TRANSFER AND EFFECTIVE GATE AREA

Let us now consider in more detail the heat-transfer coefficient and the ratio of the effective gate area to the actual area. The coefficient of heat transfer usually is found by measuring the temperature rise of a body when a known heat flux is constrained to flow through a given cross-sectional area into a liquid bath. The results are plotted in a curve of transfer coefficient versus temperature rise for a constant bath temperature. Grassman[4] has published such curves for the transfer of heat from a metallic surface to a liquid helium bath. One of his curves is reproduced in Fig. 8. Additional studies[8] of the heat transfer to a liquid helium bath are currently underway. It is obvious that the resultant h calculated in this present paper is far from the values previously determined. Three major factors must be considered when comparing the transfer coefficients: the nature of the surface over which the heat transfer takes place, the analytical results used to separate h and f, and bath-temperature variations.

The path that the heat flux must follow after leaving the gate is through layers of SiO, along and through a Pb layer, through glass to the liquid helium and from SiO to liquid helium. Therefore, the surfaces in direct contact with the helium are glass and SiO. As yet we have not made controlled experiments, such as those run by Reeber and Grassman on a metal to helium transfer, for a transfer from SiO and glass to helium. Reeber plans to investigate the transfer from various materials to liquid helium.

The second difficulty in comparing heat-transfer coefficients is that the values presented in the literature are obtained by direct measurement, whereas the values presented in this paper are determined by applying an analysis to the measurements. As pointed out before, the transfer coefficient cannot be measured directly for a gate since the heat flux does not enter the bath through the same cross-sectional area as it enters the gate. The product of the area spreading factor f and the coefficient of heat transfer h at the gate is measured directly. It is interesting to compare the product fh with Grassman's result. From Fig. 8 it can be seen that order of magnitude agreement is obtained for low-temperature differences, but the transition to a more efficient heat-transfer process occurred at a much lower temperature difference than found by Grassman.

It is to be expected that fh would be closer to the measurements made for simple flux patterns, since by using the f factor we are in effect replacing the complex pattern of heat flow by an effective isothermal area. In order to calculate the temperature distribution across a nonisothermal substrate, it is necessary to use the temperature variation of h, not of fh, and we must therefore separate the product.

Fairly good agreement between the analytical temperature distribution and the measured distributions, was obtained by using the calculated heat-transfer coefficient. It is to be noted, however, that the theoretical distribution is determined primarily by the transfer-coefficient values for low-temperature differences (< 49 m$^{\circ}$K). In this region of temperature difference, the heat-transfer coefficient is less than 22 mw/$^{\circ}$K. No value this low has been obtained in any of the metal-to-helium measurements. Theoretical temperature distributions calculated by using transfer coefficients in the range found for metal to helium do not agree at all with the experimental values.

The role of the higher heat-transfer coefficients (for temperature differences > 49 m$^{\circ}$K) is to shift the positions x_{b_1} and x_{b_2} at which the heat-transfer process changes. For a particular gate, this determines how the distribution varies with power level. The

theoretical prediction was in the right direction for all cases for which the calculation was attempted.

The third problem involved in comparison of heat-transfer coefficients is that the bath temperature is not constant during the measurement of f_h. Although the bath temperature is varied less than 200 $m^{\circ}K$ during the experiment, the effect on the heat-transfer coefficient can be quite large. Unfortunately, it is necessary to vary the bath temperature when using the gate thermal hysteresis type of measurement. One method of keeping the bath temperature fixed while determining K is to vary the width of the hysteresis loop by running transitions in the presence of an external magnetic field. This method will be incorporated into future tests. The present results, therefore, are to be interpreted as an average value of h for bath temperatures between 3.55 and 3.75$^{\circ}K$.

Another effect which makes comparison of heat-transfer coefficients difficult, although it can change h only by a factor of two, is the orientation problem: A sample located horizontally at the bottom of a container will have a higher heat-transfer coefficient than a vertically mounted sample.

At present, in view of the fairly good agreement between the measured and calculated temperature distribution, it is

recommended that the nonlinear heat-transfer coefficient h shown in Fig. 8 be used in calculations. It must be kept in mind, however, that these values have yet to be verified.

The ratio of effective-to-actual gate area was shown in Fig. 10. In the temperature distribution experiments, the power dissipated in the gate was high enough to raise the gate temperature approximately 80 $m^{o}K$. Therefore, the value of f was of the order of two or less. A plot which is useful in determining the heat-transfer coefficient at the gate for fairly high-temperature rises encountered is shown in Fig. 13. This curve emphasizes the high-temperature difference range; it is to be noted that for a given power density in this range, h and hf are fairly close. Grassman's curve is seen to give order of magnitude agreement, although his value of h does not vary at the rapid rate shown by the other curves.

Figure 10 indicates that the ratio of effective-to-actual gate area increases for narrow gates. The reason for this is that the effective area can be described as

$$A_{eff} = l\,[w + \epsilon\,(\Delta T)]\,, \qquad (14)$$

where ϵ accounts for the spreading of heat flux into the substrate.

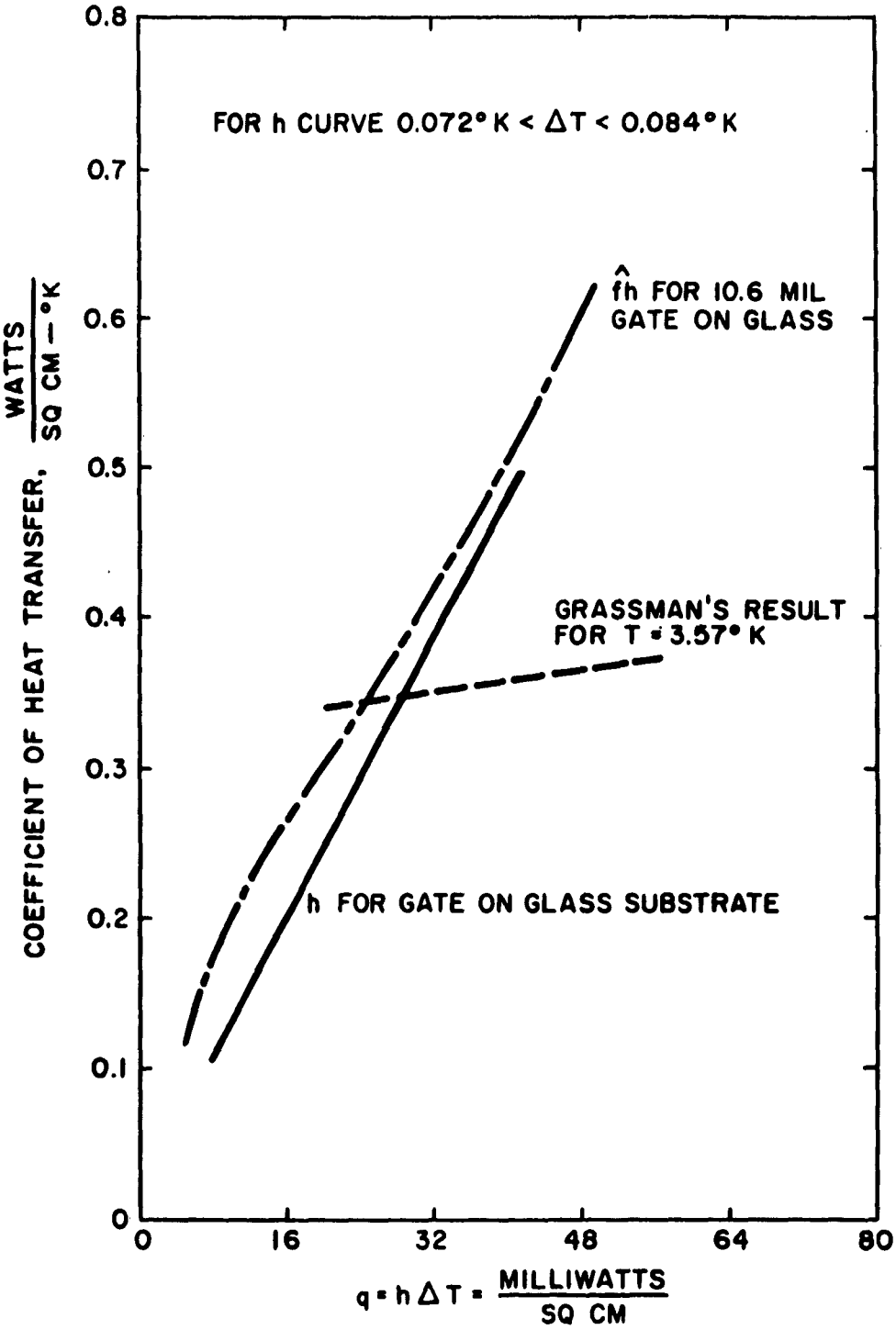


Figure 13 - Heat transfer coefficients as a function of power densities encountered in the temperature distribution runs.

Therefore, as w is increased, the effect of ϵ is decreased. This explains why, for a constant area, the gates of narrow width have a larger K. This result was shown in Fig. 5. There is a correction, similar to ϵ, for the length of the gate. However, since the gates are all relatively long, the correction is negligible. The ratio of effective to actual area approaches unity as the temperature $\Delta T(0)$ is raised.

The temperature rise of a gate can be given as

$$\Delta T = \frac{J^2 \rho_s}{h_0 \left(1 + \frac{\epsilon}{w}\right)} , \qquad (15)$$

where ρ_s is the resistivity per square of the gate and J is the density of the current flowing in the normal gate. If J represents the critical current density of the gate, then the selection of a narrow gate will minimize the temperature rise. This effect was observed during the experiments. Therefore, a "sense" gate which is usually operated with a fairly high average current should be made as narrow as possible to prevent thermal latching.

THERMAL CALCULATIONS FOR CRYOTRON SYSTEMS

An exact calculation of the behavior of a thermally coupled cryotron system is very difficult because of the nonlinear nature of the problem. It is possible, however, to investigate two extreme cases, the "loosely coupled system" (cryotrons widely spaced on a substrate with low thermal conductivity) and the "tightly coupled system" (closely spaced cryotrons placed on a substrate of very high thermal conductivity). The loosely coupled system is treated by linear methods, but the entire system is treated as an isotherm for the tightly coupled case.

Ittner [1] has considered the isothermal case and has estimated cryotron packing densities based on the dissipation resulting from the operation of multiple cryotron switching loops at very high switching rates. We will repeat Ittner's analysis for the tightly coupled case and will consider a range of operating frequencies.

The heat balance shown in Eq. (11) reduces, for the case of an isothermal substrate, to

$$P_{av} = h_0 \Delta T(0) \left[A_g + (A_s - A_g) + A_s \right]$$

$$= 2 A_s h_0 \Delta T(0). \qquad (16)$$

The maximum number of identical cryotrons, each dissipating P_{av}, that may be placed on the infinitely conducting substrate, before the temperature rise exceeds $\Delta T(0)$, is given by

$$\frac{n_\infty}{A_s} = \frac{2h_0 \Delta T(0)}{P_{av}} . \qquad (17)$$

It was shown[2] that the average power dissipated in a cryotron flip-flop, with electrical time constant τ_e, driven with a square wave at frequency f, is

$$P_{av} = \frac{I^2}{2} R f \tau_e \frac{1 - \exp - \frac{2}{f\tau_e}}{\left(1 + \exp - \frac{1}{f\tau_e}\right)^2} , \qquad (18)$$

where I is the working current and R the gate resistance. The average power dissipated when the flip-flop is driven at a duty cycle $\tau_d f$ is given by:

$$P_{av} = \frac{1}{2} I^2 R \tau_e f \frac{1 - \exp - \frac{2\tau_d}{\tau_e}}{\left(1 + \exp - \frac{\tau_d}{\tau_e}\right)^2} , \qquad (19)$$

where τ_d is the length of time that the gate is resistive.

Then following Ittner,* if we introduce the field H in oersteds associated with current I, and the gate surface resistivity ρ_s (resistivity per unit gate thickness), the average dissipation is given as

$$P_{av} = \frac{H^2 \rho_s A_g}{2(.4\pi)^2} \frac{D}{y} \frac{1 - \exp - 2y}{(1 + \exp - y)^2} , \qquad (20)$$

where $y = \frac{\tau_d}{\tau_e} = \frac{D}{f\tau_e}$; $D = \tau_d f$.

The normalized cryotron density is plotted as a function of normalized frequency, $1/y$, in Fig. 14. Note that for large values of $1/y$ (high frequency), the average power and hence the number of cryotrons tends to limit.

Typically, H may be of the order of 50 oersteds for tin, and ρ_s may be 0.02. Tables I and II illustrate some cryotron densities that are obtainable in the tightly coupled case for several maximum ambient temperature rises, gate areas, and frequency parameters. The approximate change in H associated with each temperature rise is included in the table.

---

* There is a difference of two in average power calculated from Eq. (20) for infinite frequency. This comes about because Ittner chooses a loop with four gates while Eq. (20) applies to a loop with two gates.

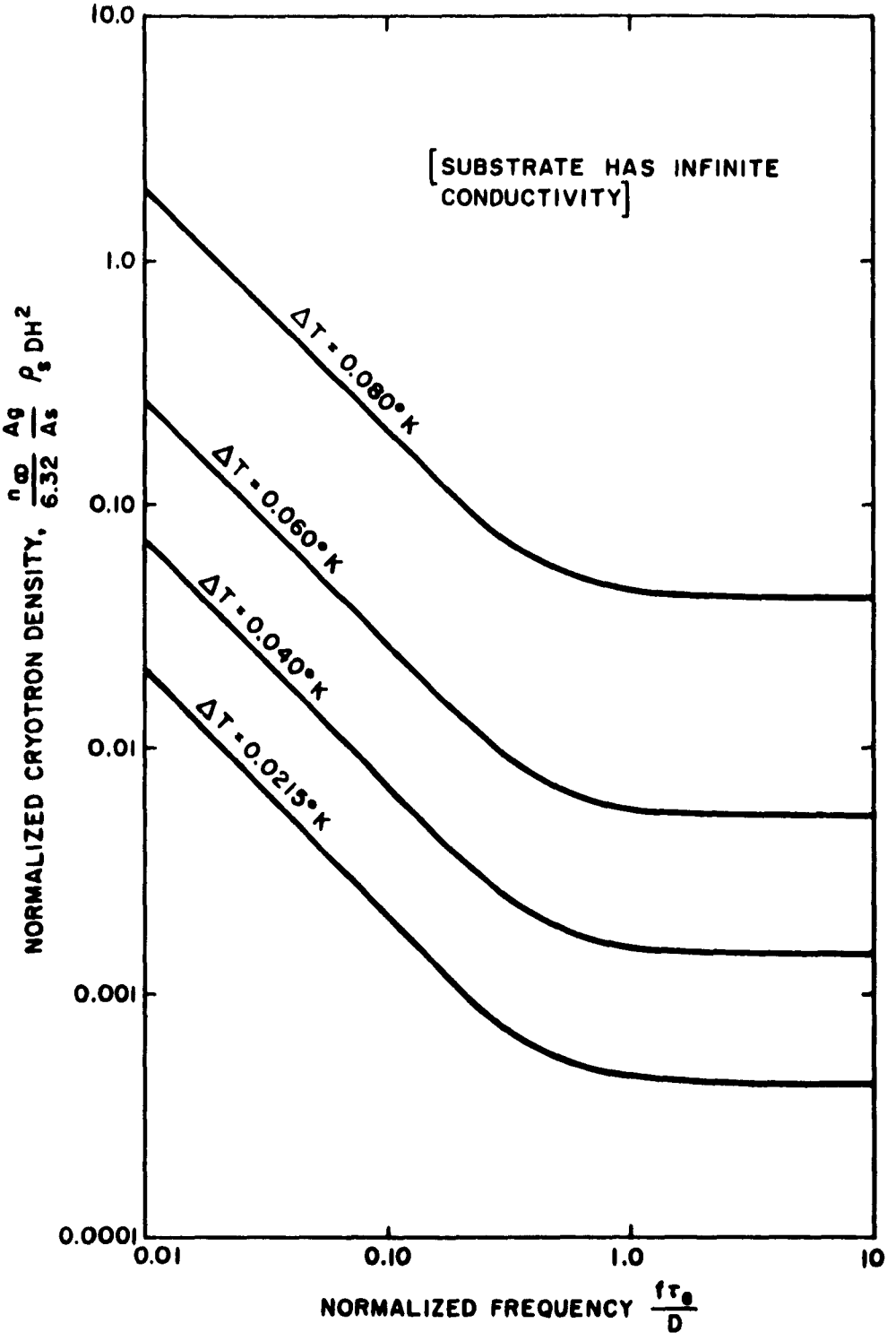


Figure 14 - Cryotron density as a function of frequency.

Table I - Cryotron Density with an Infinitely Conducting Substrate

(Gate size - 0.01 x 0.001 inch)

Frequency f, cps	Electrical time constant τ_e, sec	Duty factor D	Number of cryotrons per square inch of substrate area, n_∞/A_s			
			$\Delta T=0.0215^{\circ}K$ $\Delta H=3$ oersteds	$\Delta T=0.040^{\circ}K$ $\Delta H=6$ oersteds	$\Delta T=0.060^{\circ}K$ $\Delta H=9$ oersteds	$\Delta T=0.080^{\circ}K$ $\Delta H=12$ oersteds
10^9	10^{-9}	0.01	542	1815	6,640	51,700
10^9	10^{-9}	0.10	54.2	181.5	664	5,170
10^9	10^{-9}	0.50	11.05	37.0	136	1,059
10^8	10^{-9}	0.01	542	1815	6,640	59,700
10^8	10^{-9}	0.10	59.0	197	720	5,620
10^8	10^{-9}	0.50	27.1	90.5	333	2,590
10^8	5×10^{-9}	0.01	542	1815	6,640	51,700
10^8	5×10^{-9}	0.10	54.2	181.5	664	5,170
10^8	5×10^{-9}	0.50	11.8	39.4	144.5	1,125
10^6	10^{-9}	0.01	2710	9060	33,000	259,000*
10^6	10^{-9}	0.10	2710	9060	33,000	259,000*
10^6	10^{-9}	0.50	2710	9060	33,000	259,000*
10^6	5×10^{-9}	0.01	707	2360	8,700	68,000
10^6	5×10^{-9}	0.10	541	1000	6,800	51,600
10^6	5×10^{-9}	0.50	541	1800	6,800	51,600

* Since this number would represent an area greater than one square inch of substrate area, it is unrealistic.

Table II - Cryotron Density with an Infinitely Conducting Substrate

(Gate size - 0.10 x 0.01 inch)

Frequency f, cps	Electrical time constant T_e, sec	Duty factor D	Number of cryotrons per square inch of substrate area, n_∞/A_s			
			$\Delta T=0.215^\circ K$ $\Delta H=3$ oersteds	$\Delta T=0.040^\circ K$ $\Delta H=6$ oersteds	$\Delta T=0.060^\circ K$ $\Delta H=9$ oersteds	$\Delta T=0.080^\circ K$ $\Delta H=12$ oersteds
10^9	10^{-9}	0.01	5.42	18.15	66.4	517
10^9	10^{-9}	0.10	0.54	1.81	6.64	51.7
10^9	10^{-9}	0.5	0.11	0.37	1.36	10.6
10^8	10^{-9}	0.01	5.42	18.15	66.4	517
10^8	10^{-9}	0.10	0.59	1.97	7.20	56.2
10^8	10^{-9}	0.50	0.27	0.905	3.33	25.9
10^8	5×10^{-9}	0.01	5.42	18.15	66.4	517
10^8	5×10^{-9}	0.10	0.542	1.815	6.64	51.7
10^8	5×10^{-9}	0.50	0.118	0.394	1.45	11.25
10^6	10^{-9}	0.01	27.10	90.6	333	2590*
10^6	10^{-9}	0.10	27.10	90.6	333	2590*
10^6	10^{-9}	0.50	27.10	90.6	333	2590*
10^6	5×10^{-9}	0.01	7.07	23.6	87.0	680
10^6	5×10^{-9}	0.10	5.41	18.0	68.0	516
10^6	5×10^{-9}	0.50	5.41	18.0	68.0	516

* See note in Table I.

Several conclusions can be immediately drawn from the tables. The packing density can be appreciably increased by:

1. using small area gates,
2. designing circuits to allow for suitable operation with high ambient rise,
3. using low duty factors.

These recommendations are all fairly obvious and the tables or Fig. 14 can be used to estimate the weight of each conclusion. It is also obvious from the results that there is practically no thermal limit on the density for low-frequency operation, if moderate ambient rises are allowed.

Let us now consider the effect of using substrates, such as glass, which have a relatively poor thermal conductivity. In order to investigate this problem we will approximate the system of coupled cryotrons as a linear system. To insure some reasonable degree of accuracy, we will assume that the temperature rise of a particular cryotron, due to heating of its closest neighbor, is one tenth of the temperature rise caused by its self-dissipation. This places some severe restrictions on the problems that we can consider.

Let us replace a complicated array of cryotrons by a chain of parallel gates and assume that we can determine the temperature

rise due to the thermal coupling by imposing linearity. That is, the transfer coefficient at a particular gate is determined by the power dissipated by only that gate. Under these conditions, the temperature rise of each gate is described by the system

$$\begin{bmatrix} \Delta T_1 \\ \cdot \\ \cdot \\ \cdot \\ \Delta T_n \end{bmatrix} = \begin{bmatrix} \frac{1}{K_{11}} & \frac{1}{K_{12}} & \cdots\cdots & \frac{1}{K_{1n}} \\ \cdot & & & \\ \cdot & & & \\ \frac{1}{K_{n1}} & & \cdots\cdots & \frac{1}{K_{nn}} \end{bmatrix} \begin{bmatrix} P_1 \\ \cdot \\ \cdot \\ \cdot \\ P_n \end{bmatrix} \quad . \quad (21)$$

K_{ii} is the total "self-conductance" (previously referred to as K), K_{ij} is the "mutual conductance" between gates i and j, and P_i is the power dissipated by the i^{th} gate.

The self-conductance can be approximated by the equation

$$K_{ii} \approx h_0 A_g \left(N + \frac{M}{w} \sqrt{\frac{k\delta}{2h_0}} \right) \quad , \quad (22)$$

where M and N are functions of the temperature rise. M and N are shown in Fig. 15. It was mentioned earlier that the temperature distribution may be approximated to a first order by the form

$$\Delta T = \Delta T(0) \exp - L \sqrt{\frac{2h_0''}{k\delta}} \; x \quad , \quad (23)$$

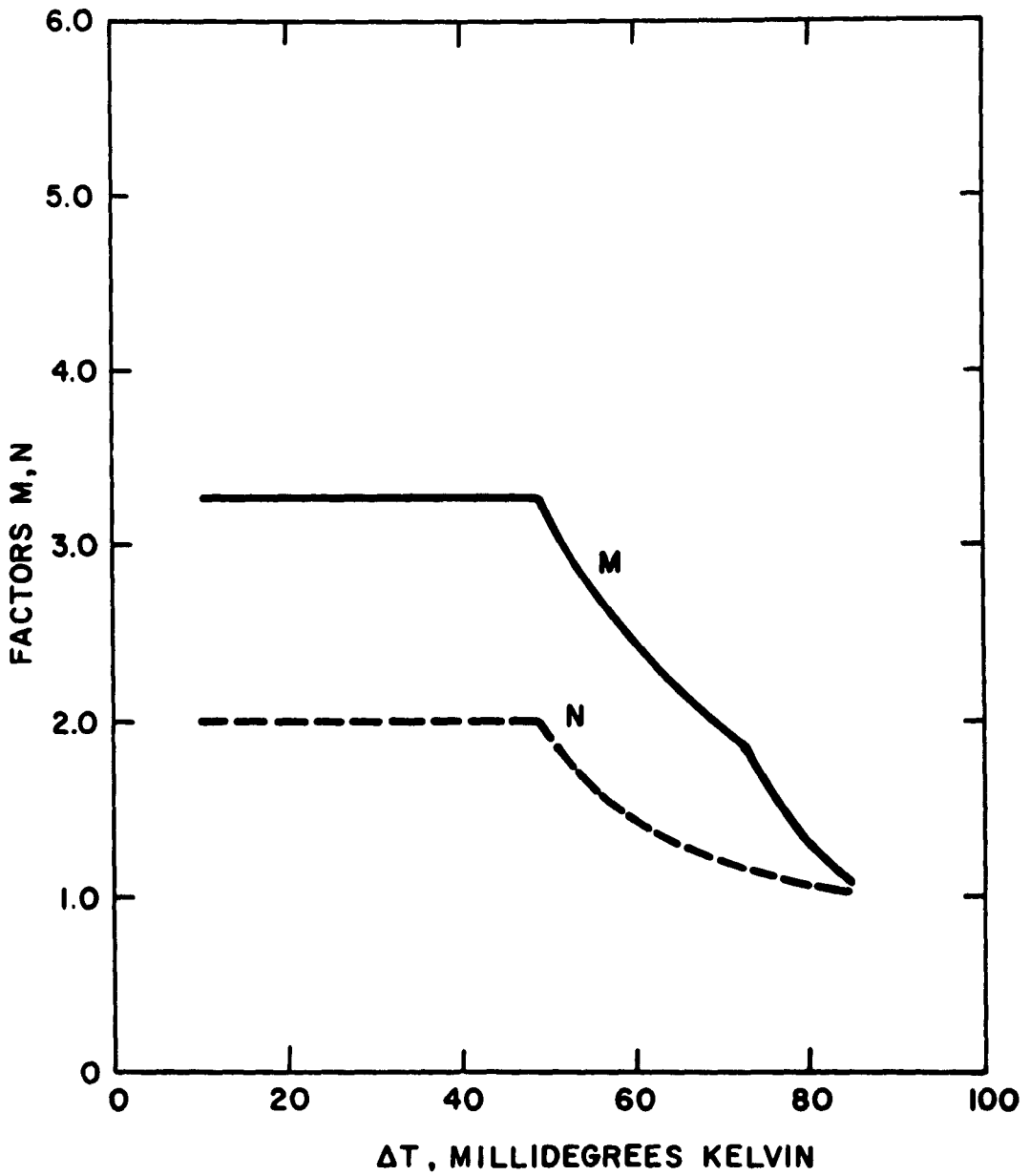


Figure 15 - M and N factors as functions of average temperature rise of each cryotron.

where L is a number less than unity. Eq. (23) may then be used to determine the mutual conductance as

$$\frac{1}{K_{ij}} = \frac{1}{K_{jj}} \exp - L\sqrt{\frac{2h_0''}{k\delta}}\, d_{ij} \quad , \tag{24}$$

where d_{ij} is the distance between the i^{th} and the j^{th} gates. Eq. (24) describes the matrix elements in Eq. (21).

If we assume that the gates are identical, equispaced, and that each is dissipating power P_{av} and has conductance K, the temperature rise of a particular gate is given as

$$\Delta T = \frac{P_{av}}{K}\left[1 + \exp - L\sqrt{\frac{2h_0''}{k\delta}}\, d + \exp - 2L\sqrt{\frac{2h_0''}{k\delta}}\, d + \ldots\ldots + \exp - nL\sqrt{\frac{2h_0''}{k\delta}}\, d\right] \quad . \tag{25}$$

For large n, Eq. (25) becomes

$$\Delta T \approx \frac{P_{av}}{K\left(1 - \exp - L\sqrt{\frac{2h_0''}{k\delta}}\, d\right)} = \frac{P_{av}}{K_{eff}} \quad . \tag{26}$$

Eq. (22), (26), and (17) and the definition

$$n + 1 \triangleq \frac{A_s}{\ell d} , \approx n , \tag{27}$$

where n is the number of gates for the loosely coupled case, can be combined to express n in terms of n_∞ as

$$\frac{n/A_s}{n_\infty/A_s} = \left(N \frac{W}{d} + \frac{M}{d} \sqrt{\frac{k\delta}{2h_0}} \right) \left(1 - \exp - L \sqrt{\frac{2h_0''}{k\delta}} d \right). \tag{28}$$

The reduction of cryotron density brought about by using a low conductivity substrate is shown in Fig. 16 as a function of substrate conductivity. The spacings indicated are chosen so that the linearity criteria will not be violated at a conductivity of 0.030 watt/^{o}K-cm.

It is possible to decrease the spacing and still have an accurate description of n/n_∞ at low conductivity. Table III shows the maximum densities that can be described by the linear theory for a conductivity of 0.001 watt/^{o}K-cm (glass).

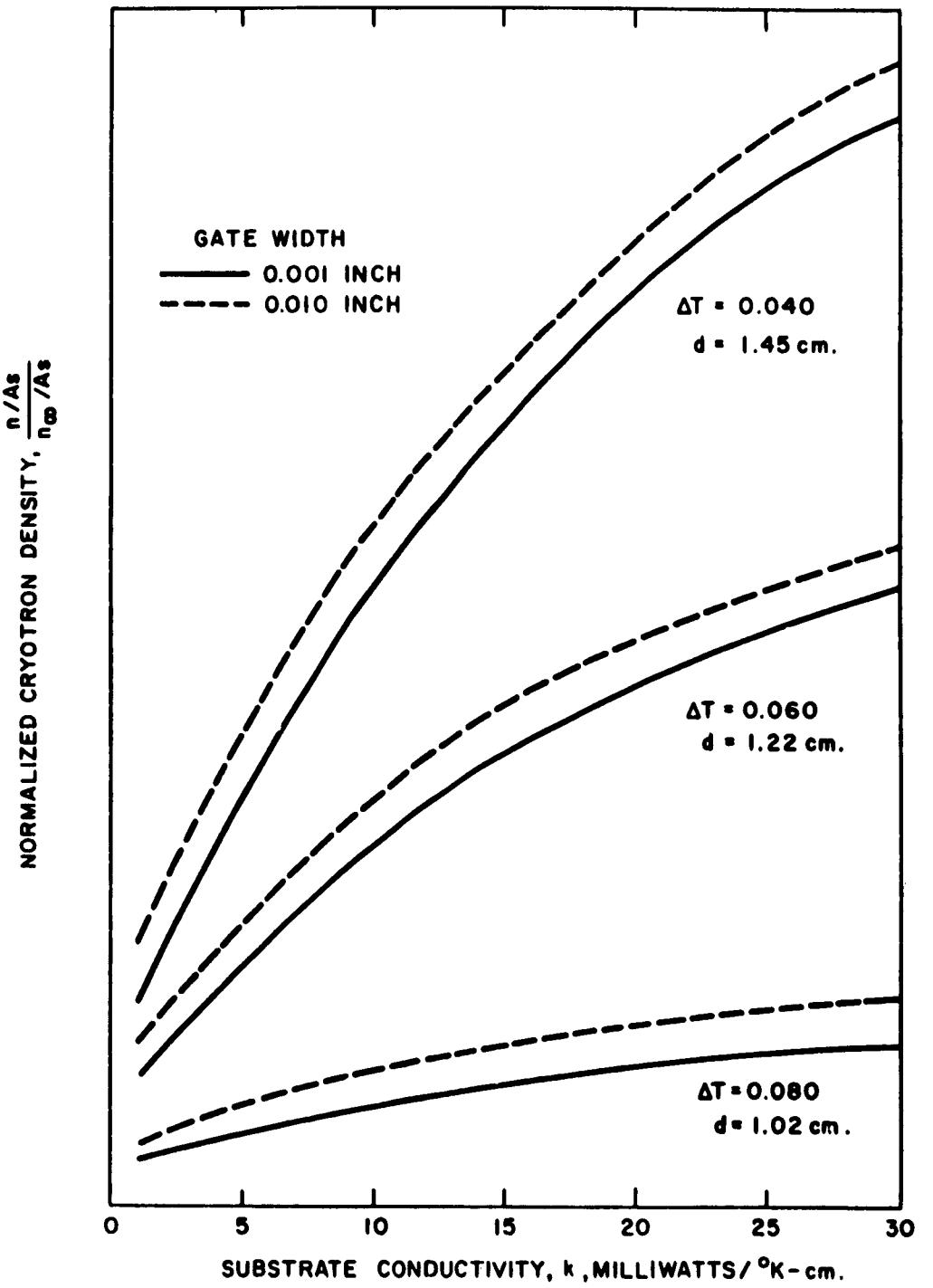


Figure 16 - Variation of cryotron density with substrate conductivity (weak coupling).

Table III - Reduction in Cryotron Density for Glass Substrate

($k = 0.001$ w/^{o}K-cm)

$\Delta T(0)$, ^{o}K	d_{min} cm	n/n_{∞} gate width = 0.001 inch
0.040	0.25	0.353
0.060	0.23	0.186
0.080	0.19	0.053

Table III indicates that with a glass substrate, it would be possible to safely use about one fifth as many cryotrons as a high-conductivity substrate would permit, allowing for moderate temperature rises, but about one twentieth if we are to allow higher increases in ambient temperature. The percent decreases as the restriction on temperature rise is relaxed, because the ultimate value, n_{∞}, increases very rapidly with ΔT.

Figure 16 shows that the cryotron density increases with increasing substrate conductivity and, for fairly wide spacing, that the number of cryotrons on a substrate with a conductivity of 30 mw/^{o}K-cm (superconducting niobium) and with a moderate temperature rise is about one fifth of the ultimate value. It

also shows that the percent of the ultimate number can be increased by using wider gates; however, as shown in Tables II and III, the decrease in the ultimate number , n_∞ would be much too severe to be of any good.

Just how high a conductivity is required to approach the ultimate number of cryotrons? To get an approximate answer to this question, let us return to Eq. (26). The effective conductance can be expressed as

$$K_{eff} = h_0 A_g N \left(1 - \exp - L\sqrt{\frac{k\delta}{2h_0''}}\ d\right) + h_0 A_g \frac{M}{w}\sqrt{\frac{k\delta}{2h}}\left(1 - \exp - L\sqrt{\frac{k\delta}{2h_0''}}\ d\right). \quad (29)$$

The second term in Eq. (29) describes what may be called the substrate component of the effective conductance. A plot of this term is shown in Fig. 17. This term clearly tends to saturate and it reaches about 95 percent of its final value when

$$\frac{1}{L}\sqrt{\frac{k\delta}{2h_0''}}\ \frac{1}{d} = 10. \quad (30)$$

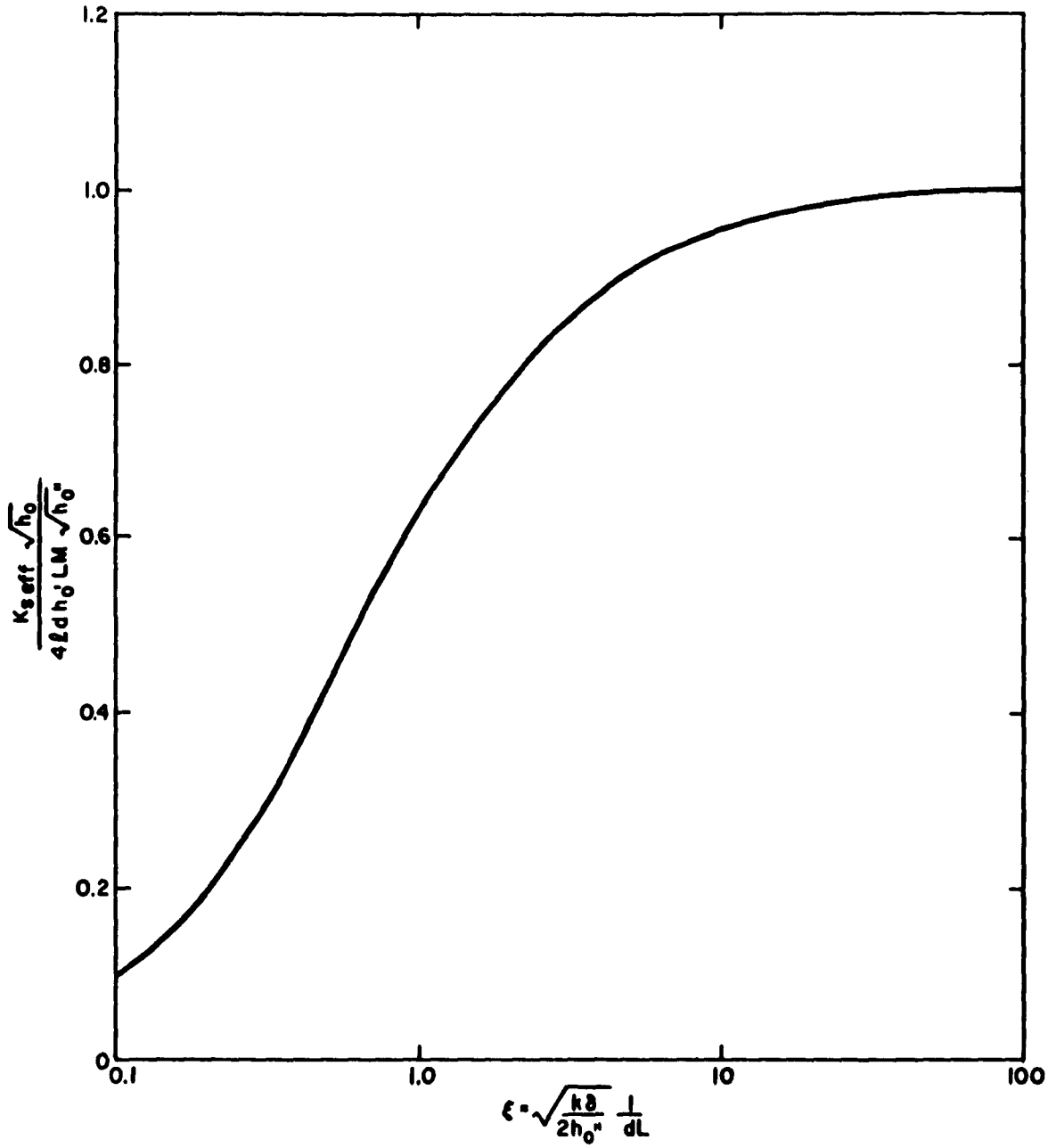


Figure 17 - Effective substrate conductance as a function of substrate conductivity.

If we assume a value of $h_0'' L^2 = 0.007$ w/°K-cm, a substrate thickness and spacing of 0.1 cm, we find that the 95 percent point occurs at a conductivity of about 0.150 w/°K-cm. A better assumption might be that the substrate temperature is near the gate temperature. Table IV lists the conductivity required to reach 95 percent of the saturation value for several temperatures.

Table IV - Approximate Values of Conductivity Required to Produce Nearly Isothermal Substrate.

(δ = d = 0.1 cm)

ΔT °K	k, w/°K-cm	Possible Material
0.0215	0.20	Al_2O_3; Pb*
0.040	0.36	Al_2O_3; Pb*
0.060	0.88	Al_2O_3; Pb*
0.080	5.00	Al, Cu

* Superconducting

Table IV lists the approximate values of k required to produce a fairly isothermal substrate and in this way to obtain cryotron densities comparable with n_∞. Note that if we operate with low or

moderate ambient temperature rise, superconducting lead would suffice as a substrate material. This is attractive, since the substrate could also serve as the ground plane. Sapphire would also suffice for a moderate temperature rise, and aluminum would be required for the high ambient rise.

CONCLUSIONS

The following conclusions are made with regard to the design of a cryotronic computer.

1. Miniaturization of cryotron dimensions contributes greatly to the realization of high packing densities. Intuitive geometric reasoning would lead one to this conclusion. Consideration must be given, however, to the dissipative effects. Fortunately, not only does the average dissipation per cryotron decrease in proportion to the cryotron area, but also the ratio of effective to actual gate area increases as the cryotron dimensions are decreased. Therefore by using small cryotrons, the packing density can be increased while the refrigeration problem is somewhat lessened. The spacing between gates is governed by the permissible ambient temperature rise.

2. Appreciable gains in packing density can be obtained by designing circuits that can operate in an environment with a fairly large ambient temperature rise. In this way, with an increase in temperature, the very nonlinear increase in the coefficient of heat transfer to the helium bath can be utilized most efficiently.

3. Substrate materials having thermal conductivities of the order of 1 watt/^{o}K-cm, or greater, at liquid helium temperatures, are required for densely packed computers. This suggests using such materials as sapphire, aluminum, or copper.

ACKNOWLEDGMENT

The author would like to thank Dr. N. H. Meyers for contributing many useful suggestions and ideas during the course of this work. The assistance of Mr. H. Martin in performing the experiments and computing results is gratefully acknowledged.

REFERENCES

1. W. B. Ittner, III, "Speed and Power Balance in Cryogenic Circuits," Project Lightning, Eighth Quarterly Progress Report, pp. 99-119.

2. H. Sobol, "Combined Electrical and Thermal Analysis of a Cryogenic Circuit," Project Lightning, Sixth Quarterly Progress Report, Vol. II, pp. 15-46.

3. Project Lightning, Eighth Quarterly Progress Report, pp. 42-55.

4. P. Grassman and A. Karagounis, Proceedings of the Fifth International Conference on Low Temperature Physics and Chemistry, Madison, Wisconsin, August 26-31, 1957, pp. 41-45.

5. Project Lightning, Sixth Quarterly Progress Report, p. 62.

6. Project Lightning, Eighth Quarterly Progress Report, pp. 42-55.

7. R. Jones, IBM Research, private communication.

8. M. Reeber, IBM Research, private communication.

APPENDIX III

The Variation of Cryotron Critical Currents as a Function of Trapped Flux in the Ground Plane

A. E. Brennemann

ABSTRACT

The dc critical currents of superconducting tin and indium films were found to be a function of the azimuthal angle of the device as it was immersed into the liquid helium. The dc critical current of one sample was reduced from its maximum value by a factor of two when the ground plane was normal to the horizontal component of the external field. It is believed the critical currents are reduced by flux trapped in the ground plane as it becomes superconducting. The critical currents are independent of orientation once the sample is immersed in the helium. This paper describes the experiment and discusses the results.

INTRODUCTION

The dc critical currents of superconducting tin and indium films, above a lead (Pb) ground plane, have not been reproducible in all experiments when the films were repeatedly immersed into a constant temperature helium bath. In some instances the critical current of a film varied as much as two to one between successive immersions. The purpose of this experiment was to observe whether a magnetic field at the surface of the helium, such as the earth's field, could be a cause of the irreproducible critical currents. Part of the experiment was to produce an external field in addition to the earth's field at the surface of the helium and observe the critical currents.

EXPERIMENTAL PROCEDURE

The dc critical currents of three samples, each with four films of superconducting tin or indium above a lead (Pb) ground plane, were measured between repeated immersions of the films into a constant temperature helium bath. The azimuthal angle of each sample was changed in angular intervals of 45 degrees between immersions for a full rotation of 360 degrees while the plane of the sample was

always vertical. The critical currents were recorded as a function of the azimuthal angle.

The first experiment was performed with only the earth's field in the region of the surface of the helium bath, and a second experiment was performed in the presence of an external field much larger than the earth's field.

The dc critical current of a film is defined for this experiment as the current that produces the smallest detectable dc voltage across the film network as it just begins to switch from the superconducting to the normal state. The smallest detectable voltage is in the order of 0.1 to 0.5 microvolt and represents resistances in the order of 10^{-5} to 10^{-6} ohms.

The results from the experiments using the above procedure are given below.

RESULTS

Figure 1 shows the dc critical currents of four indium films on a single substrate as a function of the azimuthal angle upon immersion into the helium. Each experimental point is an average of the critical currents of the four films on a substrate. The plane

of the substrate was perpendicular to the angular position shown in Fig. 1.

The results for the first experiment, in which only the earth's field was present at the helium surface, are shown as the outer curve in Fig. 1. The critical currents vary by a factor of two as a function of rotation. The angular position, at which the critical currents were a minimum, corresponds to the direction of the maximum value of the horizontal component of the earth's field.

It is believed, from this experiment, that flux is trapped in the ground plane as it is cooled below its critical temperature and thereby reduces the film critical currents. The position at which the ground plane reaches its critical temperature, about 7.2°K, was in the proximity of the surface of the helium. The amount of flux trapped was directly proportional to the magnitude of the component of magnetic field normal to the ground plane as it was cooled. The magnitude of the horizontal field component would be a function of the azimuthal angle and would be largest when the ground plane is perpendicular to the earth's field direction.

This experiment was performed on two other samples. The results for these samples were very similar to those described above,

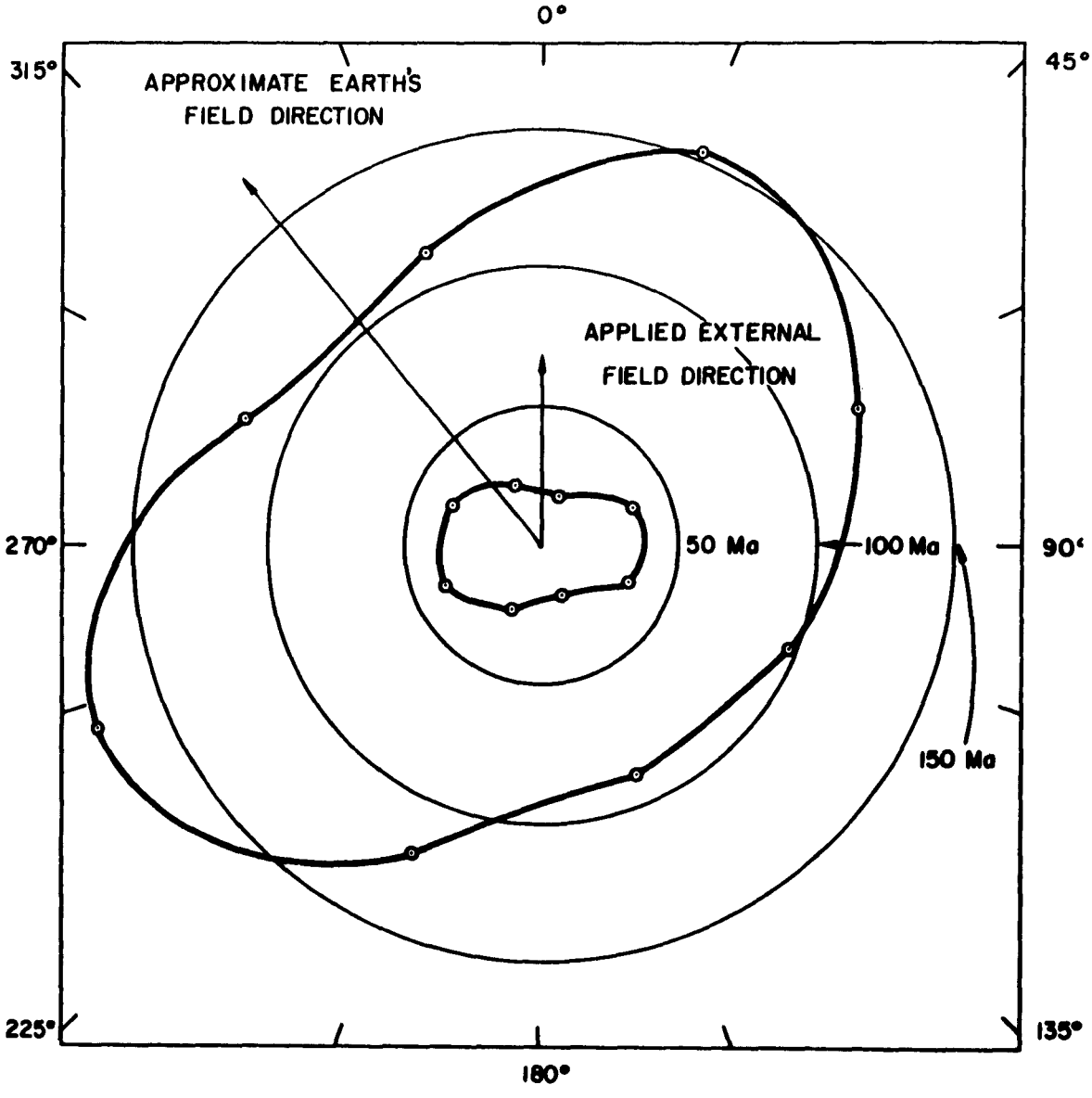


Figure 1 - The critical currents of an indium film above a ground plane as a function of azimuthal orientation as the film is immersed into the helium bath. The plane of the film is perpendicular to the angular direction shown. The scale is 50 ma per division and the reduced temperature $T/T_c = 0.89$. The film width is 0.009 inch wide and approximately 8000 Å thick.

but the variations in the critical currents were not as large as those shown in the outer curve of Fig. 1.

A second experiment was performed on the same sample as in the first case in which an external magnetic field, approximately five to ten times the magnitude of the earth's field, was applied in the vicinity of the helium surface. The inner curve of Fig. 1 represents the critical currents that were measured in the second experiment. The critical currents were reduced by roughly a factor of five and again the minimum critical currents occurred at the position at which the horizontal component of the external field is a maximum. Also, the ratio of the maximum to minimum currents is still two to one.

The results from these two experiments indicate that a field-free region is necessary at the surface of the helium bath in order for the critical currents of superconducting films above a ground plane to be independent of orientation. In one instance a field-free region was provided as the lead (Pb) ground plane was cooled through its critical temperature. The sample was lowered into the helium through the field-free region at the angular position at which the critical currents would normally have a minimum for the presence of the earth's field. The resulting currents in this angular position were equal to the maximum values obtained

earlier in the presence of the earth's field. The experiment was performed only once.

In all experiments, however, the critical currents were independent of azimuthal rotation once the sample was immersed in the helium.

The effect of the azimuthal orientation upon the critical currents was briefly mentioned by D. Dumin in a departmental report. File Memorandum, EMRc 395.

SUMMARY

The dc critical currents of superconducting tin or indium film cryotrons are reduced by the amount of flux trapped in the lead (Pb) ground plane. The flux is trapped in the ground plane in the presence of an external magnetic field normal to the ground plane as it is cooled below its critical temperature. The amount of trapped flux is a function of the azimuthal angle between the ground plane and the normal component of the external field. The critical currents can be made independent of orientation if the surface of the helium bath is in a field-free region.

APPENDIX IV

Gas Phase Oxidation of Silicon Monoxide

P. White

ABSTRACT

The equilibrium constants for the reaction of gaseous silicon monoxide with both oxygen and water vapor separately have been computed from free energy data. The values obtained have been used to determine the ratio of partial pressures of silicon monoxide and silicon dioxide which exist in the gas phase in equilibrium with various partial pressures of the oxidizing gas.

By assuming that a gas phase interaction is predominant when solid silicon monoxide is evaporated in an oxygen or a water vapor atmosphere, it is shown how these results may be used to predict the composition of the resulting film. There is reasonably good agreement between predicted values and these found by experimental analysis.

INTRODUCTION

It has been known for a long time that the physical and chemical properties of evaporated silicon monoxide layers can be altered appreciably by changing the evaporation rate, the total pressure in the system, or the source-to-substrate distance. Thus, Hass[1] has described results which show that both density and ultraviolet light transmission vary continuously with change in rate, and he has also shown that "slow deposited" silicon monoxide films took up far more oxygen on exposure to air than did "fast deposited" silicon monoxide films. Hass and Salzburg[2] concluded that only monoxide was formed at higher evaporation rates and low oxygen pressure while at lower evaporation rates and higher oxygen pressures, infrared absorption bands corresponding to silicon dioxide and higher oxides were found. More recently, Siddall[3] has shown that the electrical properties of evaporated silicon monoxide layers also vary with evaporation rate, and attributes this variation to the formation of some silicon dioxide at lower evaporation rates.

In an evaporation process, there are only two ways in which the composition of the final deposit on the substrate can be altered:

1. reaction or decomposition in the molecular beam before it reaches the substrate, and

2. reaction in the solid state after condensation.

It is a little difficult to conceive of ways in which a solid state reaction could be affected by changes in evaporation rate or by alteration of source-to-substrate distance, unless the structure of the film prior to reaction was a controlling factor. If, however, reaction took place in the molecular beam between silicon monoxide vapor and residual gas in the vacuum system, it is conceivable that alteration of the source-to-substrate distance, pressure, and evaporation rate (or molecular beam temperature) would cause a change in the composition of the molecular beam striking the substrate. This paper estimates the importance of a gas phase reaction between oxygen or water vapor and the molecular beam of silicon monoxide in controlling the composition of the final evaporated layer.

A comparatively large amount of data now exists on the thermodynamic stability of silicon monoxide solid and the rate of its disproportionation into silicon and silicon dioxide. The existence of a fourth compound, Si_2O_3, has been suggested by Cremer, et al.[6,7] The experimental data seem to imply that even if pure silicon monoxide vapor was allowed to condense on a glass slide at room temperature, the solid formed would not be pure silicon monoxide. This is in direct contradiction to the work of Hass and Salzburg and

also to some preliminary observations made in this laboratory in which a material which appeared to be pure silicon monoxide was formed by evaporation in a 10^{-7} mm vacuum. There seems to be little doubt, however, that the final composition of the condensate must bear some relation to the composition of the molecular beam striking the substrate.

The composition of the molecular beam striking the substrate is estimated by calculating the equilibrium pressures of silicon monoxide and silicon dioxide in the vapor phase when silicon monoxide reacts with oxygen or water vapor. Initially, it is assumed that no silicon vapor is emitted from a source containing either solid silicon monoxide or a mixture of silicon and silicon dioxide, as suggested by the experimental observations of Honig[8] and of Schäfer and Hörnle.[9] It is shown later in the paper that even if silicon vapor was emitted from the source, the equilibrium in the oxidation of silicon to either the monoxide or the dioxide is overwhelmingly against the existence of silicon vapor. It is also assumed that no solid phase of either silicon monoxide or silicon dioxide is formed in the molecular beam.

The calculations have been carried out in the temperature range of 1300 to 1800°K. Estimates have been made of whether or

not equilibrium conditions would be approached under typical evaporation conditions.

REACTION WITH OXYGEN

Four possible reactions with oxygen are considered:

$$SiO_{(v)} + 1/2\ O_2 = SiO_{2(v)}, \quad (1)$$

$$SiO_{2(v)} = Si_{(v)} + O_2, \quad (2)$$

$$SiO_{(v)} = Si_{(v)} + 1/2\ O_2, \quad (3)$$

$$2\ SiO_{(v)} = Si_{(v)} + SiO_{2(v)} \quad (4)$$

A free energy ΔF of any reaction is related to the equilibrium constant, K_p, by the relation $-\Delta F = RT \ln K_p$, where R is the universal gas constant and T is the absolute temperature.

Tables of free energy function of gaseous silicon monoxide have been compiled by Brewer and Edwards[4] over a temperature range of 900 to 2000°K, based on the recalculated data of Schäfer and Hörnle[9] who studied the reaction at temperatures between 1336 and 1460°K. Using these values, the free energy change for the reaction

$$SiO_{2(l)} = SiO_{(v)} + 1/2\ O_2 \quad (5)$$

has been calculated by Schick.[10] In the temperature range 2000 to 3000°K this can be written $\Delta F = 173,000 - 54.2\ T$ cal/mole. This equation has, however, been used for extrapolation beyond this lower limit. Similarly, for the reaction

$$SiO_{2(l)} = SiO_{2(v)}, \qquad (6)$$

the relation $\Delta F = 127,000 - 36.0\ T$ cal/mole is linear over the range 2000 to 4000°K, but has been used in these calculations. The two reactions (5) and (6) may be combined to yield Eq. (1); thus the free energy change for Eq. (1) can be written $\Delta F = -45,800 + 18.2\ T$ cal/mole. Using the relationship $RT \ln K_p = -\Delta F = 45,800 - 18.2\ T$, values of K_p have been computed for values of T ranging from 1300 to 1800°K and are shown in Table I.

Table I - Equilibrium Constants for Reaction

$$SiO_{(v)} + 1/2\ O_2 \rightleftharpoons SiO_{2(v)}$$

T°K	K_p
1300	5.346×10^3
1400	1.50×10^3
1500	5.00×10^2
1600	1.915×10^2
1700	8.147×10^1
1800	3.837×10^1

It is realized that extrapolation of the free energy data to 1300°K might produce an error, but it is not considered that this would be significant. The equilibrium constant of reaction (1) can be written

$$K_p = \frac{p_{SiO_{2(v)}}}{p_{SiO_{(v)}} \, p_{O_2}^{1/2}} ,$$

where p denotes the partial pressures of the various components. The equilibrium ratio of the partial pressures of silicon monoxide vapor to silicon dioxide vapor can then be calculated for various partial pressures of oxygen. Figure 1 shows how this ratio varies with oxygen pressure from 10^{-4} to 10^{-7} mm at various temperatures of the molecular beam between 1300 and 1800°K. A change occurs in the $\frac{p_{SiO_{(v)}}}{p_{SiO_{2(v)}}}$ ratio when either temperature or oxygen pressure is changed.

The equilibrium constants of reactions (2), (3), and (4) are shown in Table II.

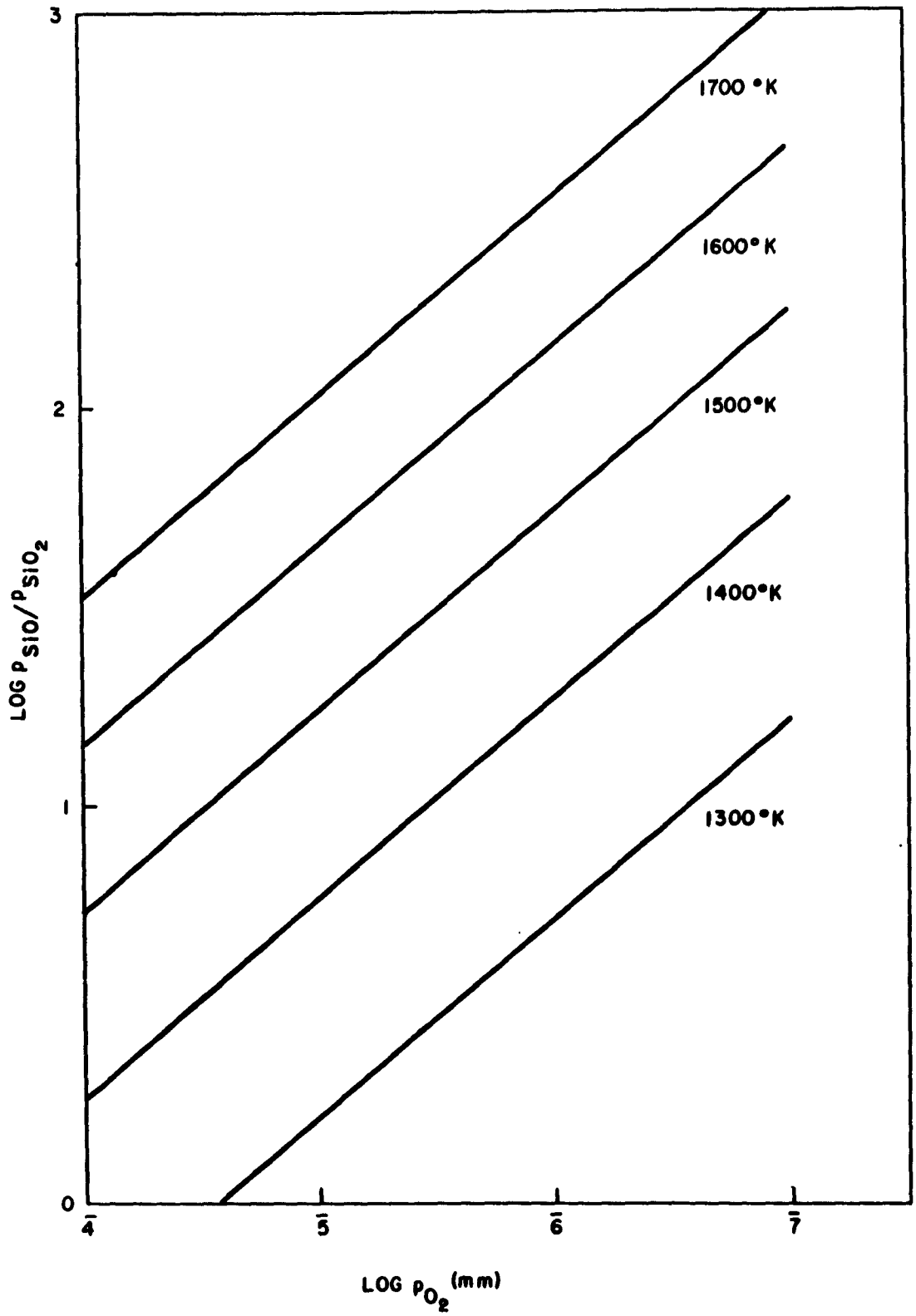


Figure 1 - Equilibrium ratio of partial pressure of silicon monoxide vapor to partial pressure of silicon dioxide vapor for the reaction $SiO + 1/2\ O_2 \rightleftharpoons SiO_2$.

Table II - Equilibrium Constants for Reactions

$$(2)\quad SiO_{2\,(v)} \rightleftharpoons Si_{(v)} + O_2$$

$$(3)\quad SiO_{(v)} \rightleftharpoons Si_{(v)} + 1/2\ O_2$$

$$(4)\quad 2\ SiO_{(v)} \rightleftharpoons Si_{(v)} + SiO_{2\,(v)}$$

$T^{\circ}K$	$K_p(2)$	$K_p(3)$	$K_p(4)$
1300	2.5×10^{-22}	1.3×10^{-18}	7.1×10^{-15}
1400	2.4×10^{-20}	3.6×10^{-17}	5.4×10^{-14}
1500	2.0×10^{-18}	1.0×10^{-15}	5.0×10^{-13}
1600	6.3×10^{-17}	1.2×10^{-14}	2.3×10^{-12}
1700	1.55×10^{-15}	1.26×10^{-13}	1.02×10^{-11}
1800	2.5×10^{-14}	9.6×10^{-13}	3.7×10^{-11}

$K_p(2)$ was computed in a similar manner to that described above for reaction (1) and was used in conjunction with $K_p(1)$ to yield equilibrium constants for reactions (3) and (4). Thus,

$$K_p(3) = K_p(1) \times K_p(2),$$

$$K_p(4) = K_p^2(1) \times K_p(2).$$

By examining the values for $K_p(2) = \dfrac{P_{Si\,(v)}\, P_{O_2}}{P_{SiO_{2(v)}}}$ and

$K_p(3) = \frac{P_{Si_{(v)}} \; P_{O_2}^{1/2}}{P_{SiO_{(v)}}}$ it would seem that under the conditions existing during an evaporation, i.e., $P_{SiO_{(v)}} \approx P_{SiO_{2(v)}} \approx 1$ mm and $P_{O_2} \approx 10^{-5}$mm, a negligible amount of silicon gas would be present in an equilibrium mixture. The original assumption that no silicon vapor is emitted from the source is now justified, since at equilibrium it would be almost completely converted to the oxide.

REACTION WITH WATER VAPOR

Only one reaction with water vapor has been considered:

$$SiO_{(v)} + H_2O = SiO_{2(v)} + H_2. \qquad (7)$$

The free energy change in reaction (7) is the algebraic sum of the free energy changes for reaction (1) quoted earlier in this paper and for the decomposition of water vapor. The latter value was obtained from tables.[11] Equilibrium constants in the temperature range 1300 to 1800°K were computed from the standard free energy changes as described earlier and are given in Table III.

Table III - Equilibrium Constants for Reaction

$$SiO_{(v)} + H_2O_{(v)} \rightleftharpoons SiO_{2(v)} + H_2$$

T^oK	K_p
1300	5.152×10^{-4}
1400	7.448×10^{-4}
1500	1.031×10^{-3}
1600	1.377×10^{-3}
1700	1.754×10^{-3}
1800	2.193×10^{-3}

The change in the ratio $\frac{P_{SiO_{2(v)}}}{P_{SiO_{(v)}}}$ with variation of the ratio $\frac{P_{H_2O}}{P_{H_2}}$ is shown in Fig. 2 for a number of different temperatures in the range normally considered.

DISCUSSION

In both Fig. 1 and 2 the ratio of $\frac{P_{SiO}}{P_{SiO_2}}$ has been computed from equilibrium constants by assuming that the partial pressures of oxygen or of water vapor and hydrogen, respectively,

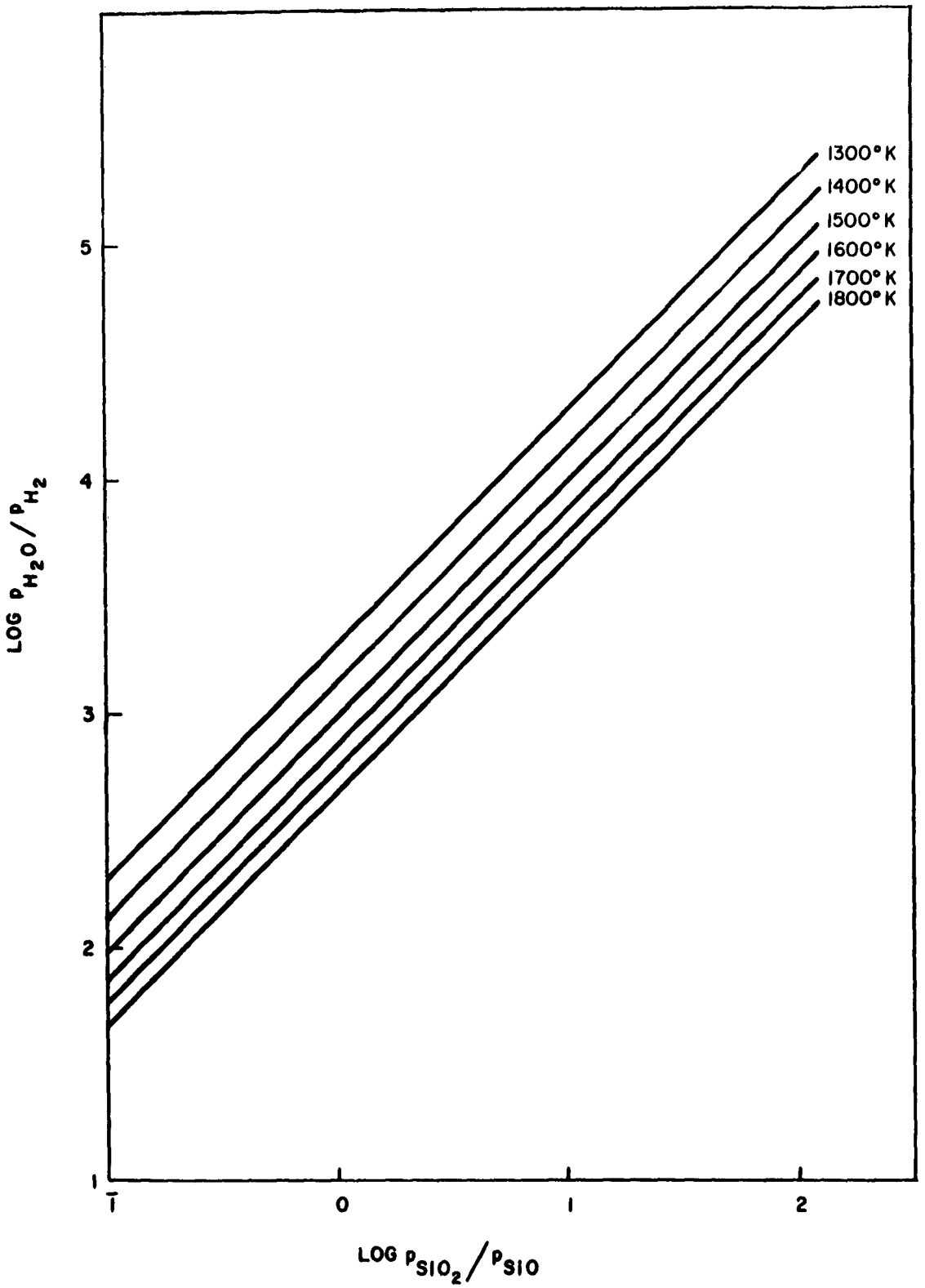


Figure 2 - Equilibrium ratio of partial pressure of silicon dioxide vapor to partial pressure of silicon monoxide vapor for the reaction $SiO + H_2O \rightleftharpoons SiO_2 + H_2$.

remain constant. Obviously, as reaction proceeds, the oxygen or water vapor is used up but the silicon oxide is continually regenerated from the source. Unless prior arrangement has been made to keep the oxygen or water vapor partial pressure at a constant value, the composition of the deposited layer would thus appear to vary continuously during the evaporation process.

It is conceivable that in an actual evaporation process the rate of reaction between silicon monoxide vapor and oxygen might be too slow to allow equilibrium to be reached in the time it takes for the molecular beam to travel between source and substrate. If this is true, the ratio of $\frac{P_{SiO_{2(v)}}}{P_{SiO_{(v)}}}$ striking the substrate would be less than the equilibrium value. Examination of the infrared spectra of the evaporated layer shows that different compositions may be obtained by the variation of evaporation parameters. There appears to be little or no dependence on whether silicon monoxide or silicon dioxide is used as starting material. Obviously some reaction is taking place with the residual gas. It is a little difficult to understand how a silicon monoxide film is produced by evaporation of silicon dioxide by any mechanism other than a gas phase decomposition.

In a bimolecular gas phase process, the rate of reaction may be written $PZe^{-E/RT}$, where Z denotes the number of

collisions per second between reacting species, $e^{-E/RT}$ the fraction of total collisions which are effective in producing reaction, E the activation energy of the reaction, and P a steric factor introduced to relate the predictions of this simple, theoretical model to experimentally observed rates. Except in some complicated reactions, P is approximately unity. In this treatment, it is assumed that P = 1 and the number of collisions between one silicon monoxide molecule and an oxygen gas molecule is determined by calculating the mean free path of the silicon monoxide molecule passing through oxygen gas, and dividing this into the distance between source and substrate. If the collision diameter for silicon monoxide/oxygen collisions is assumed to be 2×10^{-8} cm, which is about the value for most gases, the mean free path is 3.9×10^{-2} cm in a 10^{-4} mm oxygen pressure at 1500°C. If a source-to-substrate distance of 25 cm is assumed, the total number of collisions a silicon monoxide molecule makes with an oxygen molecule is 6.4×10^{2}. For equilibrium to occur under these conditions, $e^{-E/RT} = 1.3 \times 10^{-3}$, which means $E \leq 20$ k cal/mole. Unfortunately, no data exist from which we can estimate the activation energy. The only comparison we can make is with the oxidation of carbon monoxide which is a complex reaction proceeding around 1000°K with an activation energy of about 30 k cal/mole. It would not

appear too unreasonable, therefore, to expect an activation energy in the range of 20 k cal/mole for the gas phase oxidation of silicon monoxide.

To get some idea of the conditions under which equilibrium would be reached for various activation energies, the required source-to-substrate distance at $p_{O_2} = 1 \times 10^{-4}$ mm, and the required oxygen pressure with a fixed source-to-substrate distance of 25 cm have been computed for activation energies between 10 and 40 cal/mole. These are shown in Table IV.

Table IV - Geometrical Conditions Required for Complete Equilibrium in Reaction (1)

E K cal/mole	Source-to-substrate distance when pressure = 1×10^{-4} cm	Pressure when source-to-substrate distance = 25 cm
10	0.9 cm	3×10^{-6} mm
20	25 cm	1×10^{-4} mm
30	700 cm	3×10^{-3} mm
40	20,300 cm	8×10^{-2} mm

If an activation energy of 30 k cal/mole is assumed, the ratio $p_{SiO_{2(v)}} / p_{SiO_{(v)}}$ in the molecular beam when it strikes the

substrate under the standard conditions of 25 cm source-to-substrate distance and 10^{-5} mm oxygen pressure, would be 1.024. This is to be compared with a value of 5 at equilibrium.

To a large extent, the temperature of the molecular beam may be related to the rate of evaporation; hence, the relation between molecular beam composition and rate of evaporation can be observed qualitatively from Fig. 1 and 2. It should be realized, however, that the actual relationship between rate of evaporation and molecular beam temperature is dependent on source construction and geometry, and also dependent on slag formation on the surface of the material being evaporated.

It should be realized that the values shown in the figures represent equilibrium values, and the actual values may be different from these -- not because of slow reaction rates or nonequilibrium conditions, but merely by virtue of inability of each SiO molecule in the molecular beam to interact with the gas phase oxygen. This is of course true only when the SiO has to collide with an oxygen molecule in order for reaction to take place. For the decomposition of SiO_2 it is conceivable that the oxygen formed in the molecular beam is not able to diffuse out of the molecular beam and is trapped in the SiO solid. The extent to which either of these would affect the final

composition is dependent on the density of the molecular beam, i.e., the evaporation rate. A more complex situation has arisen, therefore, in which both beam temperature and evaporation rate can control the composition of the final product. Therefore, source design may be important in deciding composition.

CONCLUSIONS

Gas phase oxidation of silicon monoxide vapor seems to be a possible explanation of the large variations in experimental results observed by different workers. Presumably the partial pressures of the residual gases differ from one system to another, and the temperature of the molecular beam, which is related to source design, may also vary considerably. The results quoted in this paper indicate the composition of the molecular beam striking the substrate if complete reaction took place with an oxidizing gas in the system. The ways in which the actual results may diverge from these theoretical predictions are indicated. The values of molecular beam temperature and oxygen or water vapor pressure under which the evaporated layer is predominantly silicon dioxide or silicon monoxide are outlined.

REFERENCES

1. G. Hass, J. Am. Ceram. Soc., Vol. 3, p. 353 (1950).

2. G. Hass and C. D. Salzburg, J. Opt. Soc. Am., Vol. 44, p. 181 (1954).

3. G. Siddall, Vacuum, Vol. 9, p. 274 (1960).

4. L. Brewer and R. K. Edwards, J. Am. Chem. Soc., Vol. 58, p. 351 (1954).

5. L. Brewer and F. T. Greene, J. Phys. Chem. Solids, Vol. 2, p. 286 (1951).

6. E. Cremer, T. Kraus, and E. Ritter, Z. Elektrochem., Vol. 62, p. 939 (1958).

7. E. Cremer, A. Faessler, and H. Kramer, Naturwissenschaften, Vol. 46, p. 377 (1959).

8. R. E. Honig, J. Chem. Phys., Vol. 22, p. 1610 (1954).

9. H. Schäfer and R. Hörnle, Z. anorg. u. allgem. Chem., Vol. 263, p. 261 (1950).

10. H. Schick, Chem. Revs., Vol. 60, p. 331 (1960).

11. "Selected Values of Properties of Hydrocarbons," Circular C 461, Natl. Bur. Standards.

APPENDIX V

A Method for Detecting Imperfections in Thin Insulating Films

E. M. DaSilva and P. White

ABSTRACT

A simple apparatus has been assembled in which discontinuities occurring in insulating films may be observed in a nondestructive manner. The signal produced by scanning a metal-insulator surface with an electron probe can be used to detect a metal-insulator boundary and also to indicate variations in insulator thickness in the 100 to 1500 Å range.

The usefulness of this instrument in a study of nucleation of molecular beams of insulating materials on metal surfaces has been determined by examining test slides containing various patterns of insulators on a lead surface.

While the resolution of the instrument is limited by the radius of the electron beam used, the size of an imperfection which could be detected by the instrument is really dependent only on the current density in the electron beam. Imperfections having an area 1/100 that of the electron beam have been detected without difficulty.

INTRODUCTION

When evaporated insulating films are used as dielectric layers, the problem of shorts in the dielectric usually has to be considered. In most cases, it is not possible to observe through an optical microscope a surface discontinuity or imperfection which can be related to a shorted area. At present there is no satisfactory nondestructive method of testing insulating films prior to deposition of the upper electrode. The electron microscope undoubtedly has sufficient resolution to detect small imperfections, but this technique has two major disadvantages; only small areas can be sampled at one time; and, since surface replicas of the insulating film have to be made, this is a destructive test. A further method has been suggested[1] in which a magnetic sense coil is used to detect the field associated with a pinhole short. However, this method has limited applicability.

A technique described in this paper has none of the disadvantages outlined above. The apparatus used can be constructed simply and can, if required, be installed in the evaporator used for preparing insulating films. Using this method, it is possible to make nondestructive tests of an insulating layer prior to deposition of the upper metal electrode. The unit, which is similar in principle to the Monoscope,[2] uses an electron probe to scan the insulator surface. It

is shown that the difference in signal when the probe scans a metal-insulator interface is sufficient to permit detection of that interface. In addition, under appropriate conditions, the signal varies with insulator thickness so that the thickness differences can be detected. The construction of a simple model is described and the resolution of this model estimated by examining test slides containing patterns of silicon monoxide evaporated onto a lead film. The results are discussed and extrapolated to indicate the ultimate resolution which can be obtained from this type of instrument. It is shown that by the use of a more sophisticated electron gun this technique could be made sufficiently sensitive to detect imperfections in the insulating film of about 16,000 Å diameter. The technique is not capable of diffe - entiating between a group of holes and a thin area unless the minimum spacing between holes is an appreciable percentage of the beam diameter.

EXPERIMENTAL PROCEDURE

The experimental arrangement is shown in Fig. 1. A type 5C electrostatic deflection gun similar to that used in a Tektronix 512 oscilloscope was sealed into a glass tube. Two aquadag electrodes were painted on the inside of the tube to act as a final accelerating

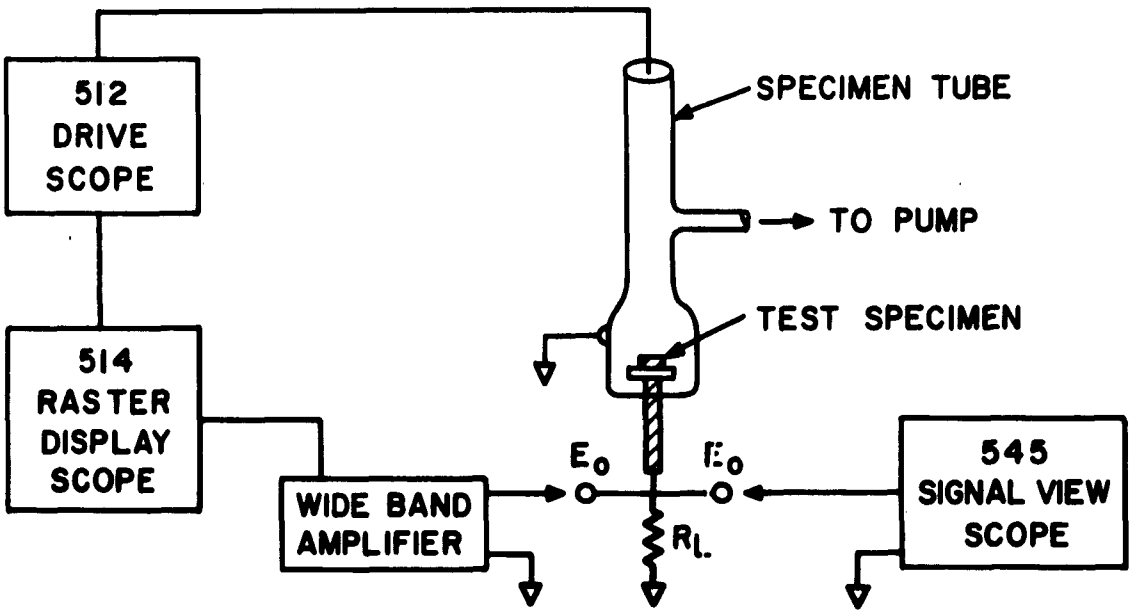


Figure 1 - Experimental layout.

anode and a secondary collector. A rectangular copper block supporting the test slide was sealed into the other end of the glass tube so that the test slide was at right angles to the electron beam. The tube was evacuated by a well-trapped mercury diffusion pumped system, and pressures in the 10^{-7} mm range were attainable without difficulty.

A low-voltage power supply was used to activate the cathode and to regulate the cathode emission in the specimen gun. Control of the electron beam was achieved by connecting the specimen gun in parallel with the gun in a Tektronix 512 oscilloscope.

The test specimens consisted of definite patterns of silicon monoxide evaporated onto a lead film supported on a glass microscope slide. These were clamped to the copper block by phosphor bronze clips which also provided a conducting path from the front of the slide to the copper block. The signal produced when the electron probe scanned the sample surface was then available by making an external connection to the copper block, and was observed on a Tektronix 545A oscilloscope which had a band width of 10 Mc.

Two modes of signal display were used: the signal response accompanying a line scan of a particular area, and a raster display in which the signal was used to modulate the intensity of the presentation. In practice, the raster display was used only to locate the position of any one area on the test slide relative to a fixed point on the drive scope tube. A line sweep could then be made on that area and the signal response observed on a viewing scope. The distance between two peaks on the display signal could be magnified by adjusting the ratio of sweep speeds of the display scope and the drive scope so that resolution of discontinuities in the test slide was not limited by resolution in the display sweep.

The secondary emission ratio is dependent on the accelerating anode potential, V_p, which has a critical value at which the

secondary emission ratio is a maximum. Figure 2 shows some representative curves[3] for lead and silicon monoxide as a function of V_p. These curves can only be used as a guide to indicate a range of operating potentials, since the secondary emission ratio is strongly dependent upon the surface conditions. The anode potential of the specimen tube was varied through the range

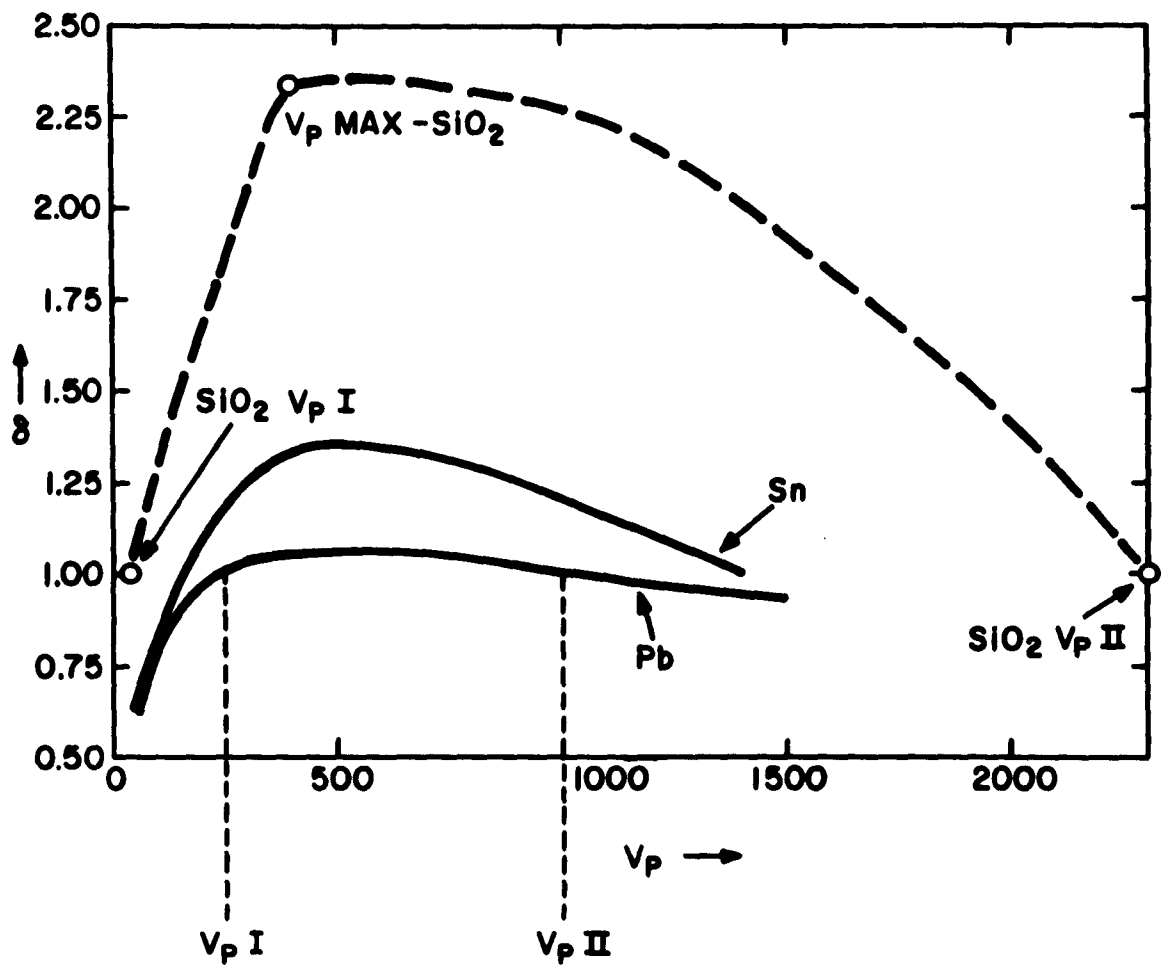


Figure 2 - Secondary emission characteristics of lead, tin, and quartz, as given in Reference 3.

indicated in Fig. 2. It was found that the best signal response occurred when operating between 1200 and 1500 volts.

The copper holder could be biased above or below the collector voltage. This could then act as a collector or emitter depending on bias voltage and the conditions of greatest signal response selected.

RESULTS

To identify the different areas on the specimen slide, it was necessary that the signal response caused by the incidence of the electron beam on those areas vary sufficiently to define and clearly delineate the areas. The first test specimen contained a pattern of large and small areas of both lead and silicon monoxide as shown in Fig. 3a. A raster display of this specimen is shown in Fig. 3b, in which the dark areas are the insulating films. The smallest area of silicon monoxide was 0.062 inch wide, which was approximately twice the beam diameter of 0.030 inch. The good agreement between Fig. 3a and 3b shows that this method can be used to identify silicon monoxide-lead boundaries.

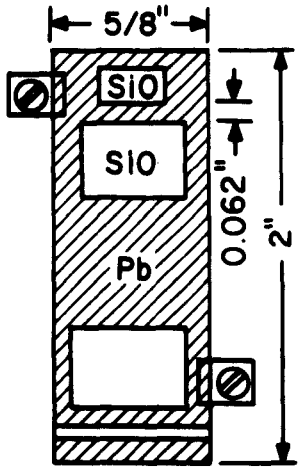


Figure 3a - Test specimen No. 1 containing lead and silicon monoxide.

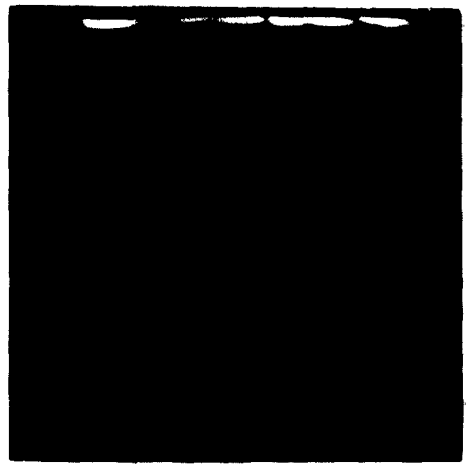


Figure 3b - Raster display of specimen No. 1.

Pensak[4] has shown that a conduction current can be induced in a thin insulating film by an electron beam, and also that the magnitude of the induced current depends on film thickness, accelerating potential, and target bias voltage. If the accelerating potential and bias voltage are kept unchanged, it should, therefore, be possible to distinguish between two insulating films of different thicknesses, if the difference in conduction current in the two films is sufficiently large.

A second specimen slide (Fig. 4a) consisted of six separate areas of insulation with thicknesses varying between 400 and 3500 Å, and one area in which the thickness varied in discrete steps from 400 to 3500 Å. Figure 4b shows a raster display of this

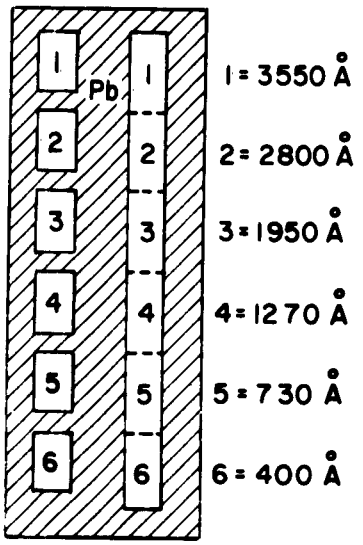


Figure 4a - Specimen of insulating areas of varying thickness.

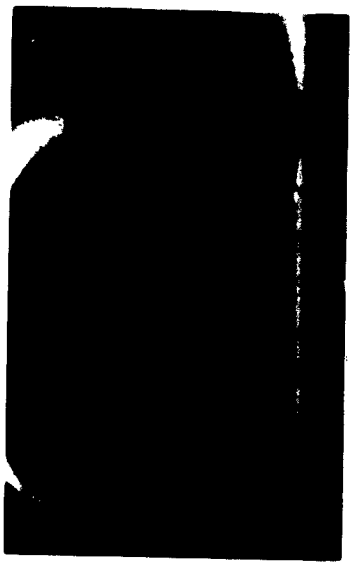


Figure 4b - Raster display showing contrast due to thickness variation

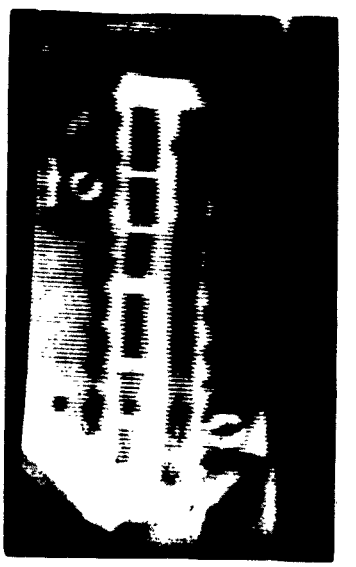


Figure 4c - Raster display in which thickness variation is not evident.

specimen. Once again, the dark areas represent the insulator, but by controlling the amplification of the signal modulating the intensity of the scope, the thinner layers can be made to appear as dark areas. It appears possible, therefore, to obtain a thickness contour picture of a thin insulation film by controlling the amplification of the signal modulating the intensity of the raster display. Figure 4c shows another raster display of the second test slide using a different signal amplification. It should be noted that for a given accelerating potential and bias voltage, there is a limited thickness range in which the variation of conduction current with thickness is sufficient to permit detection. This thickness range can be changed, however, within limits, by variation of either accelerating potential or bias voltage.

It has been shown so far that this simple apparatus is capable of detecting metal-insulator boundaries and also of detecting variations in film thickness in a thin insulator. The ability to observe a thickness variation is as useful as the detection of discontinuities or holes, since a thin spot in an insulating film is a potential source of trouble in a device.

It is important to estimate the sensitivity of this technique for detecting either discontinuities in insulators or extreme thickness variations before building a more sophisticated arrangement. In

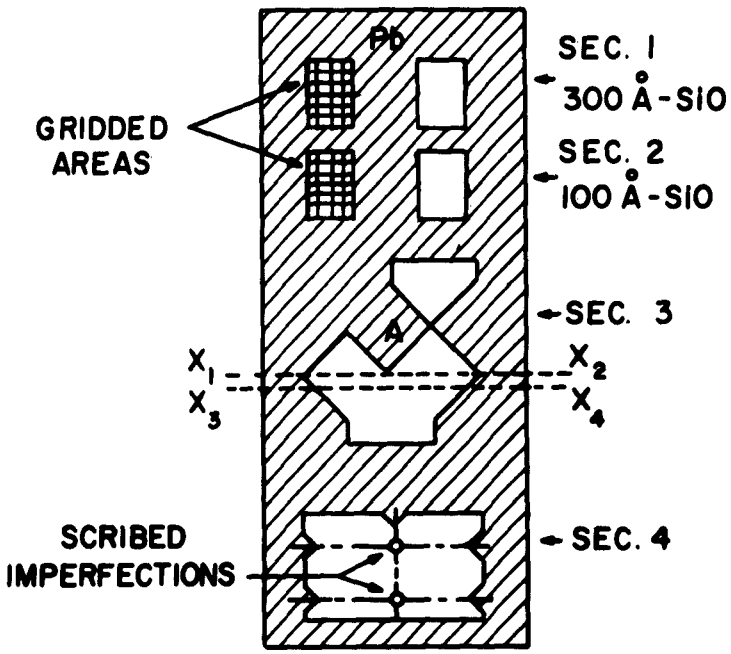


Figure 5a - Specimen No. 3, made of solid and gridded layers of varying thickness.

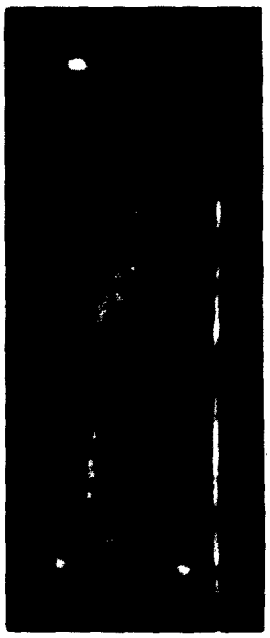


Figure 5b - Raster display of No. 3.

order to do this a test slide was fabricated as shown in Fig. 5a. A raster display is shown in Fig. 5b. The slide contained six different patterns of silicon monoxide evaporated on a lead underlayer. Two of these, marked section 1 in Fig. 5a, were 300 Å thick, one a continuous layer and the other deposited through a 144-mesh wire grid. Section 2 contained two areas of similar design to those in section 1, but the silicon monoxide layer was 100 Å thick, so that signal differences between 100 and 300 Å films could be observed. The geometrical design shown in section 3 was evaporated to provide an estimate of the minimum area of

metal which could be detected by the edge of the electron probe. The last section, 4, was a test pattern designed so that deliberate mechanical discontinuities could be introduced in defined positions. Both sections 3 and 4 were 300 Å thick films.

The signal generated as the probe scans the continuous area and the area masked by the mesh is shown for the 300 Å and 100 Å films in Fig. 6a and 6b. It is obvious that there is insufficient sensitivity to resolve the metal lines in the masked areas, but this is hardly surprising in view of the relative dimensions of the gridded area and the electron beam diameter. The gridded section contained metal lines 0.0015 inch wide separated by silicon monoxide areas 0.0075 inch wide, compared to the electron beam diameter of 0.030 inch.

For both thicknesses it appears that the resulting signal from the masked area is composed of separate contributions from the metal and the insulator sections since the magnitude of the signal generated when the probe scans the gridded section is part way between the separate signals for all lead or all silicon monoxide. It is conceivable that such a composite signal could be used to yield information on the relative areas covered by insulator and by metal, but the assumptions involved in calibration would seem to make this a rather tenuous method.

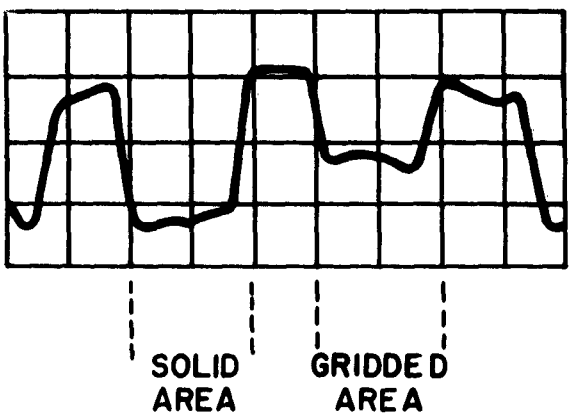


Figure 6a - Signal display for the 300 Å solid and gridded blocks, sec. 1, Fig. 5a.

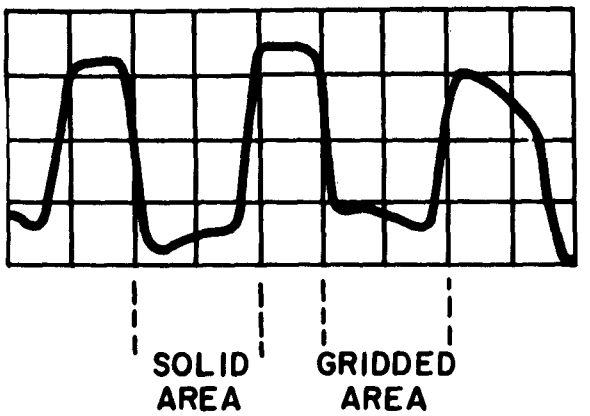


Figure 6b - Signal display for the 100 Å solid and gridded blocks, sec. 2, Fig. 5a.

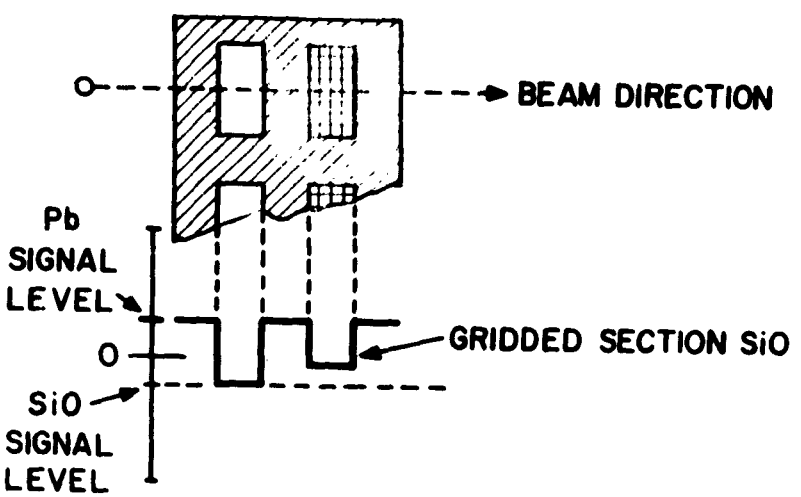


Figure 6c - Signal display associated with a beam path over the solid and gridded areas of the SiO films.

(For Fig. 6a, 6b, and 6c, probe speed = 0.01 sec/cm; view speed = 0.002 sec/cm; sensitivity = 1 v/cm.)

It should be noted that the signal difference between the gridded and continuous areas in the 100 Å films is greater than that in the 300 Å films. This is presumably due to the larger signal observed on the 100 Å continuous silicon monoxide film as shown in Fig. 6c.

In the experiment with section 3, the line of sweep was set along a known length $x_1 x_2$ (Fig. 5a). This measured the magnification which was set at about 100. The line of sweep was then lowered toward $x_3 x_4$ until the signal due to the metal film at A just disappeared. By comparing the lengths $x_1 x_2$ and $x_3 x_4$, it was possible to determine the area of the triangular metal film causing the signal. The minimum detectable area by this method was between 0.000025 and 0.00003 square inch, which is approximately 1/30 of the probe area. This technique used the extreme edge of the electron beam for sampling, and since the current density of the beam follows a Gaussian distribution this figure may be considered an upper limit.

Section 4 (Fig. 5a) contained two scratches deliberately placed so that their positions could be accurately defined with respect to the total area, as shown in the diagram. In this way, the signal associated with the inserted imperfections could be located and identified. The area of the larger of these two discontinuities was

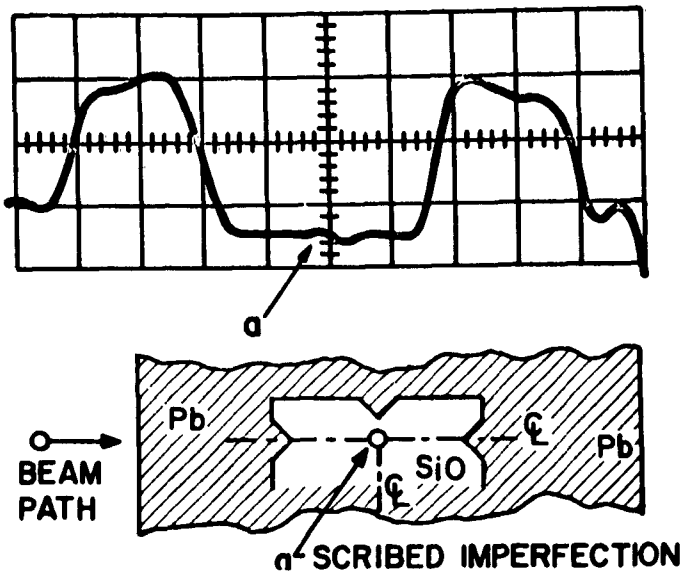


Figure 7 - Signal display for scratch located in sec. 4, Fig. 5a, and illustration of scribed imperfection.

less than 1/100 of the beam area, but, as shown in Fig. 7, it showed up as a signal difference when scanned by the electron beam.

DISCUSSION

The experimental results show that when an electron beam scans an insulator or a metal surface, sufficiently large differences in signal current exist to distinguish clearly between the two. The observed signal when an electron beam scans an insulating surface is due partly to secondary emission from the surface and partly to conduction through the film induced by the electron beam.[2] It may

be observed from the results quoted in Pensak's paper that at the lower accelerating potentials the conduction current is greater in thinner films. This may explain the observed signal difference between 100 and 300 Å insulating films.

This method makes it possible to locate holes and thickness variations in thin insulating films without destroying the film. More detailed examination of the shorted area may then be made with an electron microscope. The resolution of the instrument, defined as the smallest distance apart at which two imperfections may still be detected separately, is obviously limited by the beam diameter. The over-all sensitivity of the instrument in detecting an isolated discontinuity in a metal or insulator layer seems to be dependent only on current density in the electron beam and sensitivity of the detector. With the instrument described in this paper it has been shown that an imperfection 1/100 the diameter of the electron beam can be detected. It is now of interest to examine the lowest limit of detection which would be possible by using an electron gun with a smaller beam diameter.

In general, the signal response from a given area is directly proportional to the current in the incident beam. It would be possible to increase the beam current density and therefore the sensitivity by

using a tungsten cathode. This has a relatively low emission efficiency in comparison with the oxide coated cathode at the same temperature. It can, however, be operated at a higher temperature, 2500 to 3000°C, yielding a higher density beam. This would increase the complexity of the system and for this reason was not attempted. The minimum usable signal level is limited by the noise level associated with the electronics of the system. If the shot noise is neglected, the signal-to-noise ratio may be written as

$$S = \frac{i_p \, R_L^{\frac{1}{2}}}{(4kT\Delta f)^{\frac{1}{2}}},$$

where Δf is the band width and i_p the beam current. For the condition where the secondary emission is approximately unity (and we neglect conduction current) the signal current, i_s, is equal to the beam current. If it is assumed that a signal-to-noise ratio of ten is sufficient to observe a ten percent change in secondary emission ratio[5] the above equation can be solved to give $i_p = 0.025\ \mu a$ for values of $R_L = 10,000\Omega$ (and $\Delta f = 4$ Mc). An electron gun very similar to the one used in this experiment with a beam diameter of 0.001 inch can be obtained commercially. For a current density, J_o, of 1 amp/cm^2 which is in the

range of that produced by an oxide-coated cathode,[4] the minimum theoretical detectable area would be

$$A = \frac{i_p}{J_o} = 2.5 \times 10^{-8}.$$

This is equivalent to a circle of diameter 16,000 Å, which is approximately 1/10 of the beam diameter. It should be noted that it has been possible to observe experimentally imperfections 1/100 of the beam diameter. The resolution of the apparatus can be improved by use of an electron gun with a beam diameter less than 0.001 inch. Such guns are commercially available but these are, in general, guns in which magnetic deflection of the electron beam is used. Once again a compromise had to be made between resolution and complexity of the instrument, so the guns with the smallest diameter beams were not used.

The most sensitive form of the arrangement described above is, of course, the electron scanning microscope, with which the practical resolution is about 500 Å.[6]

REFERENCES

1. Project Lightning, Sixth Quarterly Progress Report, p. 71.

2. C. E. Burnett, "The Monoscope," RCA Review, Vol. 2, pp. 414-420 (April 1938).

3. H. Bruining, "Physics and Applications of Secondary Electron Emission," (McGraw-Hill Book Co., Inc., New York, 1954), Chap. 3.

4. L. Pensak, "Conductivity Induced by Electron Bombardment in Thin Insulating Films," Phys. Rev., Vol. 72, pp. 472-479 (1949).

5. V. K. Zworykin, J. Hillier, and R. L. Snyder, "A Scanning Electron Microscope," Bull. of the Am. Soc. for the Testing of Materials, Vol. 117, pp. 15-23 (1942).

6. D. McMullan, "An Improved Scanning Electron Microscope for Opaque Specimens," Proc. Inst. Elec. Engrs., Vol. 100, pp. 245-259 (1953).

GENERAL REFERENCES

M. Knoll and B. Kazan, "Storage Tubes and Their Basic Principles," (John Wiley and Sons, Inc., New York, 1952).

V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier, and A. W. Vance, "Electron Optics and the Electron Microscope," (John Wiley and Sons, Inc., New York, 1945).

APPENDIX VI

Simulation of the Subtractor-Ring Circuit

W. C. Carter and J. L. Sanborn

ABSTRACT

The subtractor-ring circuit was simulated by using the RL network-analyzer program. The switching speed of the tree circuits was seen to depend upon their previous state because of induced back currents. These back currents and their effect on switching times are shown graphically. Because of these effects on the subtractor, the network should be operated as a clocked device at about 500 kc, or as a free-running device with the ring cycling in about 2 μsec. Determining exact operating times would entail simulation through all data states. It should be noted that future circuits would contain in-line cryotrons having sharper characteristics which would greatly increase the operating speeds.

INTRODUCTION

The subtractor-ring circuit shown in Figure 1 and described in a previous report[1] has been simulated as a free-running network to:

1. determine the speed of network operation as the input current is varied,
2. determine the speeds at which the circuit will operate when driven by external pulses, and
3. examine the magnitude of the network currents as a function of input current and time.

SHORT DESCRIPTION OF THE SUBTRACTOR-RING CIRCUIT

The subtractor-ring circuit is basically a combination of the data-transfer circuit,[2] the four-stage clocked ring,[3] and the subtractor,[4] all described in previous Lightning reports.

When the subtractor-ring circuit is internally operated, it functions as an asynchronous (free-running) parallel counter. When current is flowing in one side of the first stage, the minuend is set equal to the contents of the accumulator by cryotrons 400 to 407. Then when the state of the first stage has been transmitted through the other three stages of the ring, by cryotrons 200 to

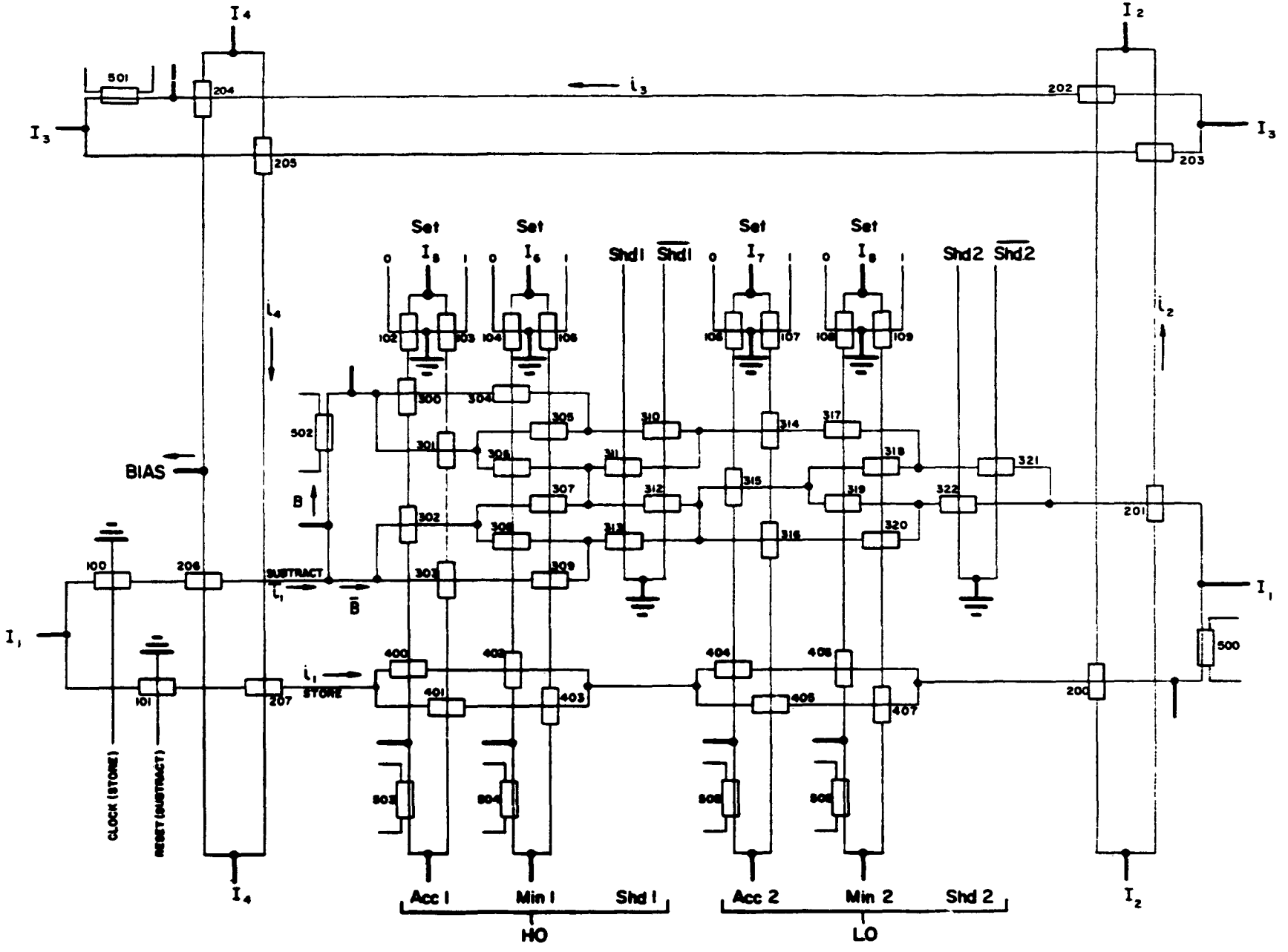


Figure 1 - A subtractor-ring circuit.

205, the first stage is reset by cryotron 207. This supplies current to the subtractor, and the accumulator is set equal to the difference between the contents of the minuend and the contents of the subtrahend by cryotrons 300 to 321. When this state of the first stage has been transmitted through the other three stages of the ring, the first stage is set to its original condition by cryotron 206.

The circuit may be operated as a synchronous parallel counter by "biasing out" cryotron 206, so that resistance is never introduced in the circuit by this cryotron. The store and subtract sequence is then triggered by the clock pulse, and the circuit will come to rest when both functions have been performed. Note that under clocked operation the duration of the subtract signal is determined by the interval between pulses.

INITIAL CONDITIONS FOR THE SIMULATION

The circuit was divided into 32 meshes, consisting of 14 resistive elements and 101 inductive elements. The assumed dimensions of the network are:

line width	0.018 inch
line thickness (Pb)	4000 Å
line thickness (Sn)	6000 Å
penetration depth (Pb)	500 Å
penetration depth (Sn)	1250 Å
crossed-film cryotron control width	0.003 inch
insulation thickness	4000 Å
in-line cryotron gate width	0.003 inch
in-line cryotron control width	0.009 inch
in-line cryotron length	0.125 inch
operating temperature	95% of critical temperature

Using the model of cryotron operation shown in Fig. 2 and described in the seventh quarterly Lightning report,[5] the characteristics assumed for cryotron operation are:

I_{crit}	300 ma
ρ	0.2375
d'	100 ma
R_{max}	0.87 m-ohm

The distribution of currents will be described by referring to the notation of Fig. 1, and giving their state as logical functions where possible. During all simulation runs, Shd 1 = 0, Shd 2 = 1. The initial conditions were:

Acc 1	= 1
Acc 2	= 1
Min 1	= 0
Min 2	= 0
i_1	= 0
cryotron 206	resistive

Thus the borrow current, B, initially goes through cryotrons 100, 206, 502, 301, 306, 311, 314, 317, 321, 201. The quantities Acc 1, Acc 2, Min 1, Min 2, i_1, B were graphed on the on-line plotter.

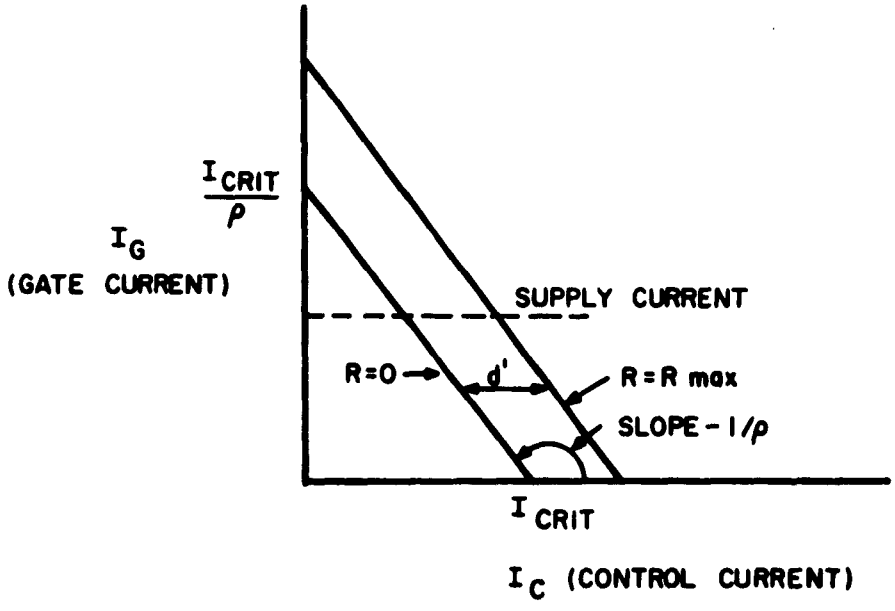


Figure 2 - Assumed cryotron characteristics.

FIRST SIMULATION

For the first simulated run, all supply currents were assumed to be 400 ma. In this case the cycle time of the ring was between 0.76 and 0.77 μsec. In the first cycle (a store operation), Min 1 and Min 2 were switched to 308 ma and 328 ma, respectively. The borrow current, B, stopped temporarily at 192 ma in the first store cycle because Min 2 was changing from 0 to 1, so that neither side could introduce resistance into the $\overline{B}$ leg. As $\overline{i}_1$ approached 0 (84 ma), B continued to decay.

Back currents were induced through the loops defined by cryotrons 502, 301, 306, 307, 302; through 302, 308, 309, 303; and through 301, 305, 304, 300. The magnitude of the currents through 303, 301, 300, 302, B, $\overline{B}$, $\overline{i}_1$ is shown in Fig. 3. The -185 ma current through $\overline{B}$ is surprising.

When current paths are switched, the effect of these currents is to slow down switching and, even worse, make switching speed dependent directly upon existing current configuration. The second cycle was to have been a subtraction cycle, with the following operation: 11 - 01 = 10.

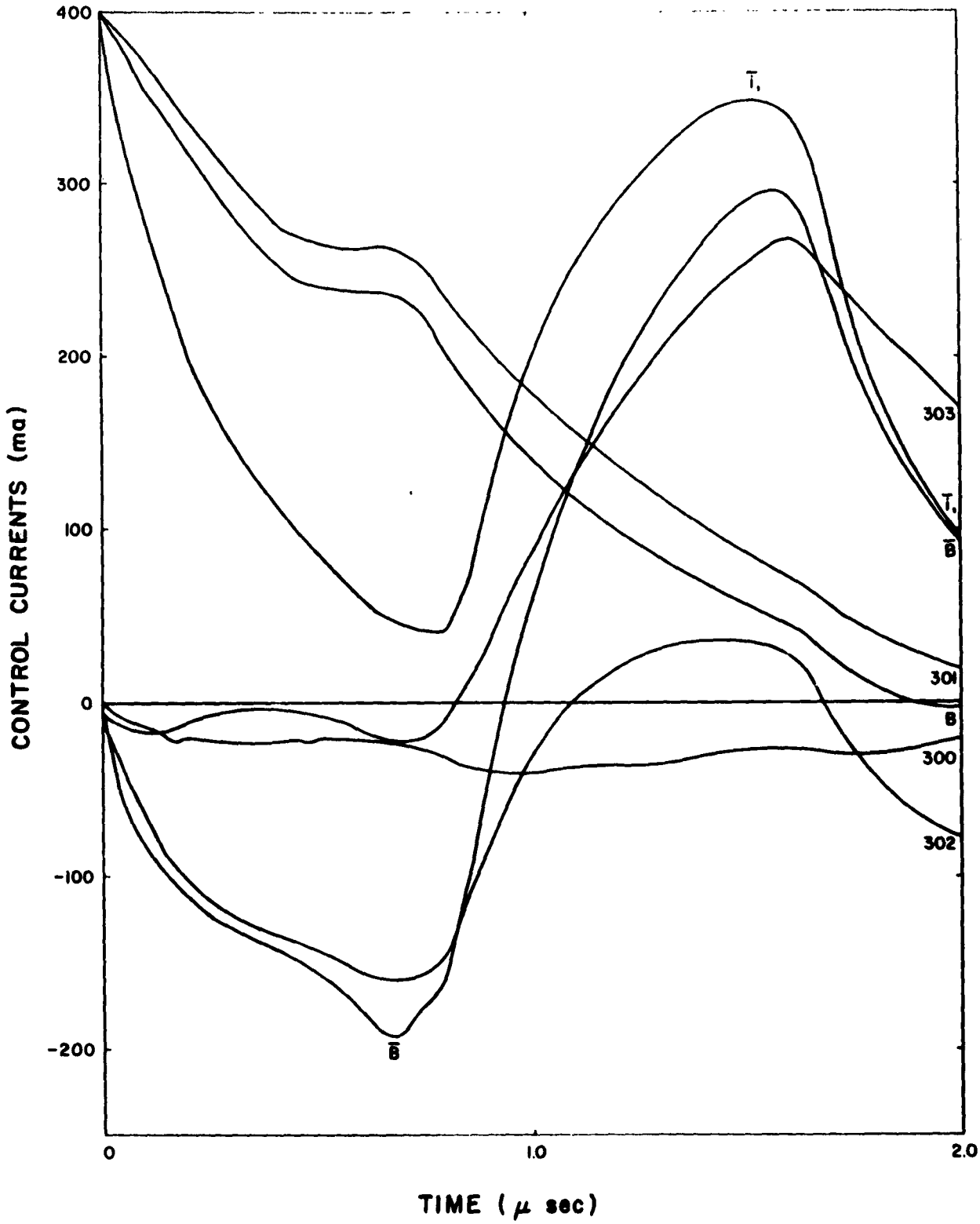


Figure 3 - Cryotron control currents as a function of time.

Min 2 was switched, however, to 204 ma instead of 0 ma, and thus current splitting (204 and 196 ma) occurred and correct circuit operation stopped.

SECOND SIMULATION

In the second simulation run, the supply currents I_1, I_2, I_3, I_4 of the ring were reduced to 375 ma, with the others remaining at 400 ma. The other initial conditions were unchanged. In this case, the cycle time of the ring was between 0.93 and 0.94 μsec. In the first cycle, Min 1 was switched to 340 ma and Min 2 to 360 ma. The borrow current fell to 136 ma, then continued to 0. In the next cycle, subtraction was correctly performed with Acc 1 remaining at 400 ma and Acc 2 falling to 84 ma value. In the next cycle, Min 1 increased to 376 ma while Min 2 correctly switched to 84 ma. In the next subtraction cycle, however, there was not sufficient time for the proper operation: 10 - 01 = 01 to be performed. Because the effect of the divided currents was of interest, the simulation was continued and the magnitude of Acc 1, Acc 2, and currents through cryotrons 300, 301, and 316 are shown in Fig. 4 and 5. The minimum current (205 ma) necessary

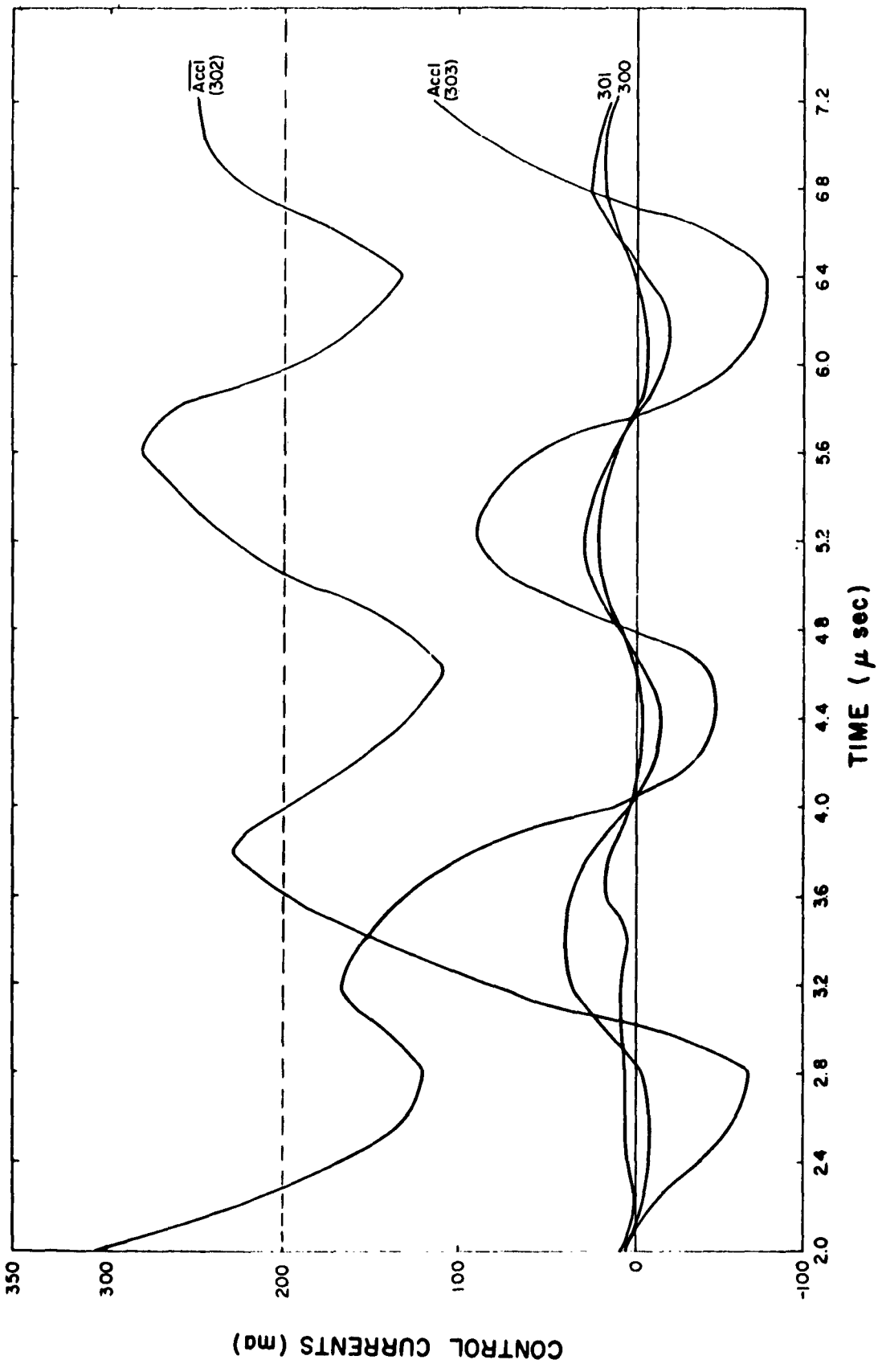


Figure 4 - Cryotron control currents as a function of time.

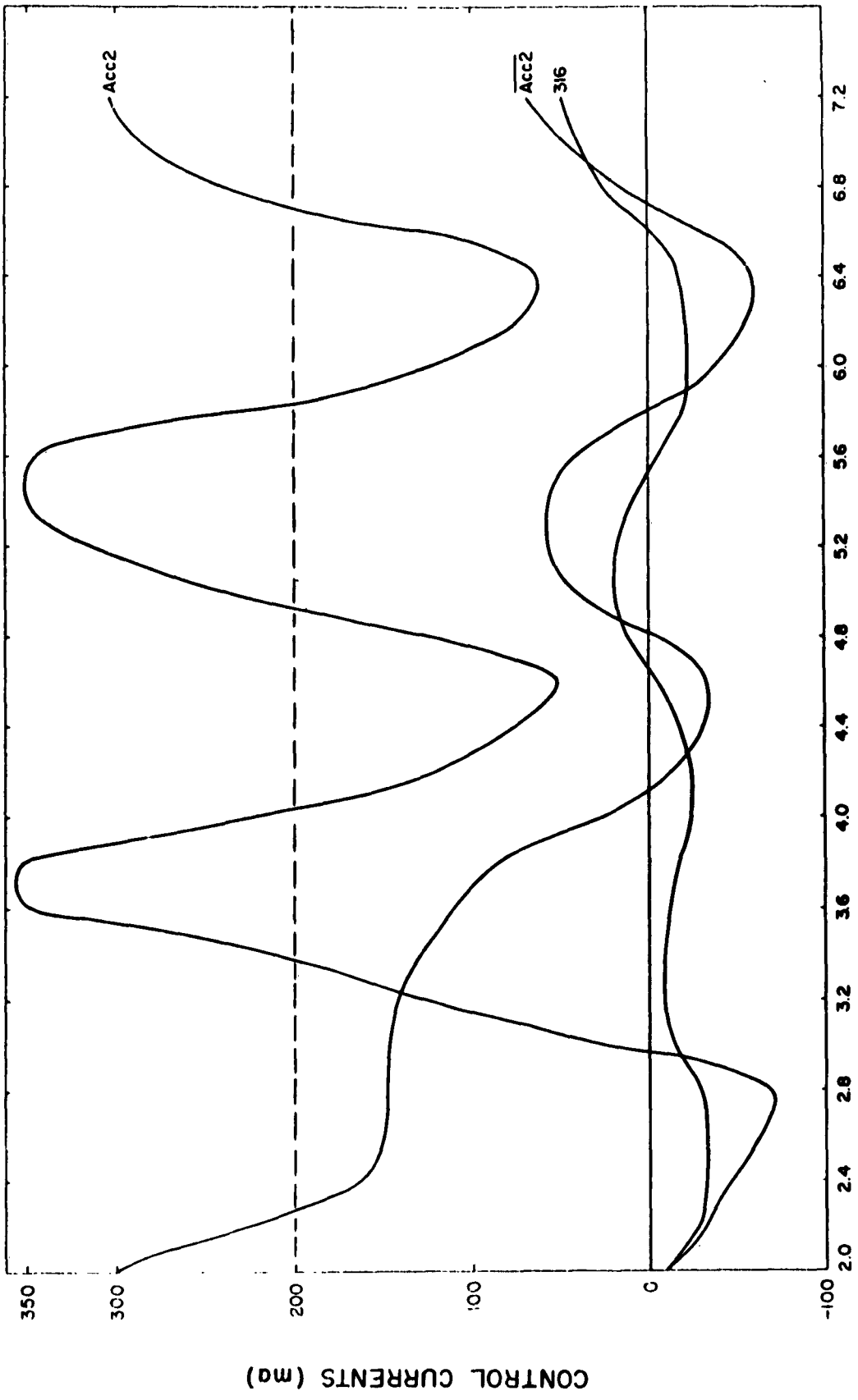


Figure 5 - Cryotron control currents as a function of time.

to induce resistivity in a cryotron is shown as a dashed line.

These graphs show the effect of back currents on cryotron current switching. The switching speed of these tree circuits clearly depends upon their previous state.

Note that in the simulation 0.76 μsec was sufficient to transfer the contents of the Accumulator to the Minuend. However, the delay between clock pulses will have to be greater than 1 μsec to allow back currents in the subtractor to die down enough for proper operation. To obtain this time would entail a simulation through all operations using all data states.

CONCLUSIONS

A. Circuit Evaluation

1. The circuit will operate as a clocked ring.
2. The circuit will operate in a free-running mode if the ring loop is slowed down sufficiently.

B. Tree Circuit Evaluation

1. The simple loops (2 path) performing storage worked well.
2. The switching speed of tree circuits depends upon their

previous state, in particular upon

a) the degree of completion of previous cryotron switching and

b) induced back currents.

3. For efficient operation of complicated tree circuits balanced branches are necessary.

REFERENCES

1. Project Lightning, Seventh Quarterly Progress Report, p. 52.

2. Project Lightning, Fourth Quarterly Progress Report, p. 49; Fifth Quarterly Progress Report, p. 49.

3. Project Lightning, Sixth Quarterly Progress Report, p. 83.

4. Project Lightning, Fifth Quarterly Progress Report, p. 53.

5. Project Lightning, Seventh Quarterly Progress Report, p. 59.

APPENDIX VII

Sequential Cryotron Switching Circuits

J. L. Rosenfeld

ABSTRACT

In order to design sequential switching circuits, the designer must have available both a device that can perform combinational logic and a device with a memory or a delay. Cryotron circuits possess both combinational logical properties and memory properties. It is the aim of this appendix to demonstrate how sequential cryotron switching circuits that make full use of the features of cryotrons can be designed.

INTRODUCTION

When the memory inherent in cryotron circuits is fully exploited, a simple sequential circuit generally results from the synthesis procedure. Furthermore, neglect of the intrinsic memory can lead to serious errors in the design of logical cryotron circuits. In this appendix a direct procedure is used to synthesize a sequential circuit from the description of desired circuit operation. Next it is demonstrated how synthesis by means of a transition matrix yields a simpler circuit. This latter synthesis procedure takes full advantage of built-in cryotron memory. The necessary restrictions on the design of circuits via the transition matrix are then discussed. Finally the hazard problem is investigated.

The logical properties of cryotrons are well known.[1,2,3,4] In this appendix, we assume that any switching function can be realized with cryotrons, and we ignore the means by which the logic is performed. It may be performed with crossed-film or with in-line cryotrons, and in one or several stages. The fact that a path becomes resistive when a logical function of certain variables equals "1" is indicated by a rectangular box in the path, crossed by lines that represent the variables. An expression for the function is written in the box. In Fig. 1e, for example, the top left box, A,

X_1X_2 : 00	01	11	10	OUTPUT Z
①	2	–	5	0
1	②	4	–	0
–	2	③	5	1
–	2	④	5	0
1	–	3	⑤	0

Figure 1a - Flow table.

X_1X_2 : 00	01	11	10
①	2	③	⑤
1	②	④	5

Figure 1b - Merged flow table.

y \ X_1X_2	00	01	11	10
0	0	0	1	0
1	0	0	0	0

Z

Figure 1c - Output matrix.

y \ X_1X_2	00	01	11	10
0	0	1	0	0
1	0	1	1	0

Y

Figure 1d - Excitation matrix.

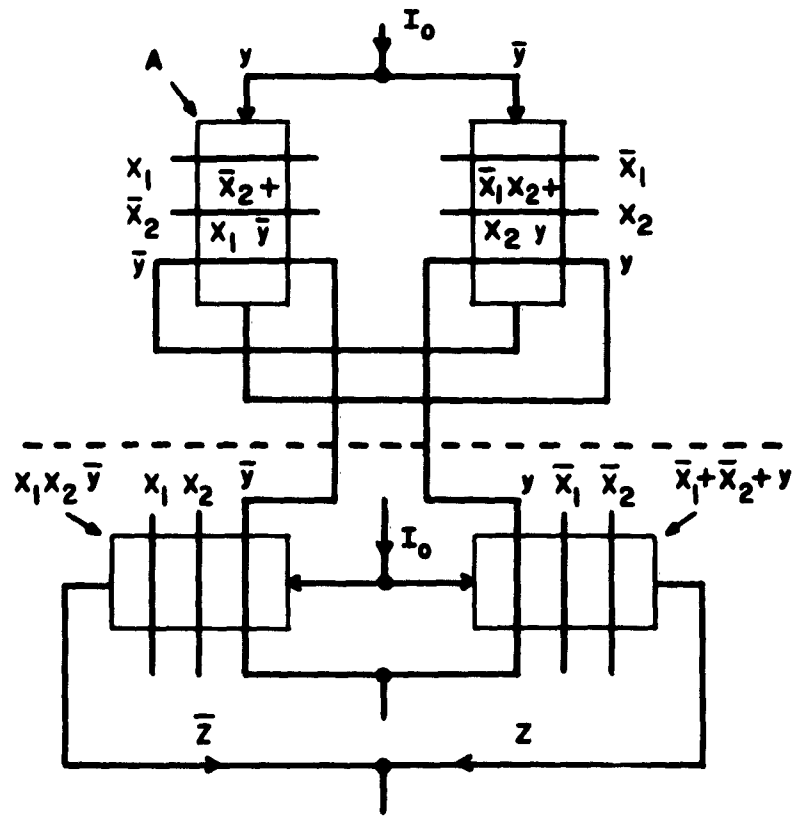


Figure 1e - Resultant circuit from direct design procedure.

represents any cryotron configuration that inserts resistance in the left-hand branch if $x_2 = 0$ (no current flows in path x_2), or if $x_1 = 1$ and $y = 0$ (current flows in path x_1 and no current flows in path y), and which forms a superconductive path otherwise.

A particular example will be helpful in illustrating general principles discussed later. Consider this description of the desired behavior of a cryotron circuit with two inputs, x_1 and x_2: only one input can change at a time, and it is assumed that the inputs change only after all circuit transients have died out; the output of the circuit is to be 1 when x_1 and x_2 are both 1, but only if x_1 is 1 before x_2; i.e., the input pair $x_1, x_2 = 1, 0$ must precede 1, 1. Several tables must be constructed during the synthesis procedure.[2,5] Figure 1a is a flow table that describes the circuit action. The circled entries represent stable states of the circuit. Uncircled entries in the table indicate the next state the circuit must assume when the input condition is changed from that input condition associated with a stable state in the same row. Dashes imply unallowed input changes. Entries in the output column are the outputs Z associated with the stable states in the same row. Figure 1b is a merged flow table corresponding to the primitive flow table of Fig. 1a. The rows of the merged flow table contain the same information as the rows of the primitive flow table after certain assignments are made for transitions corresponding to

unallowed input changes. Figure 1c is the output matrix, which contains essentially the same information as the output column of Fig. 1a. One secondary variable y is necessary in order that the circuit can assume the two states which correspond to the two rows of the merged flow table. An excitation matrix is shown in Fig. 1d. Each matrix entry Y represents the next state secondary variable y must assume for the present input conditions and state of the circuit, as represented by the column and row of the entry. This follows from the merged flow table.

In order to synthesize a circuit that performs according to the preceding specification, it is necessary to select some way of realizing the secondary variable y. For this purpose, an elementary cryotron circuit, like that of Fig. 2, is chosen. The condition y = 1 represents the state of the circuit when $i = I_o$, or all the current flows down the left-hand branch. Conversely, y = 0 represents the case in which i = 0, or all current flows down the right-hand branch.

DIRECT SYNTHESIS

First we shall perform the synthesis by a direct procedure. The excitation matrix of Fig. 1d indicates that current should be forced to flow in the left-hand branch of the secondary variable when

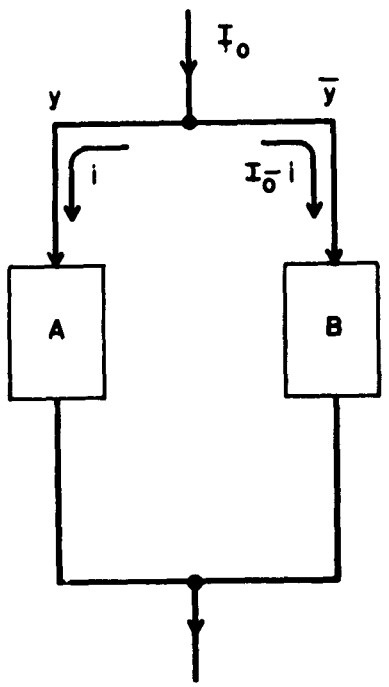


Figure 2 - Elementary cryotron circuit.

$Y = 1$, or $\bar{x}_1 x_2 + x_2 y = 1$. Current must be forced to flow in the right-hand branch when $Y = 0$, or $\bar{x}_2 + x_1 \bar{y} = 1$. These conditions can be effected by making all paths in the right-hand branch resistive when $Y = 1$ and causing at least one path to be superconductive when $Y = 0$. Similarly, all paths in the left-hand branch must be resistive when $Y = 0$, and a superconductive path must exist otherwise. The upper half of Fig. 1e represents the circuit for secondary variable y. Notice the feedback (cross-latching) that exists in this circuit. The lower half of Fig. 1e represents a realization of the output matrix of Fig. 1c. The reader can easily verify that this circuit operates according to the preceding description of desired behavior.

This direct synthesis procedure is easily generalized to any sequential circuit problem with any number of secondary variables. One elementary cryotron circuit like that of Fig. 2 is used to represent each secondary variable. The branches of these elementary circuits are made resistive according to the entries in the excitation matrices for the secondary variables. The direct procedure is described by Caldwell[2] (pp. 608-613), and Fig. 1e is equivalent to his Fig. 14-35.

SYNTHESIS BY USE OF TRANSITION MATRIX

Cryotronic sequential circuit synthesis with the aid of the transition matrix is based on the fact that once current is established in a branch, it continues to flow in that branch until the branch becomes resistive. That is, when current flows in a superconducting path of a cryotron circuit, the current persists in that path until some cryotron in the path becomes normal. (The effects of mutual inductance have been ignored in this appendix. Currents can be induced in superconducting loops, and existing persistent currents can be diminished because of the mutual inductance; nevertheless, the magnitude of these effects is so strongly dependent upon the type

of cryotron circuitry used that it is impossible to make any general statements about the consequences of mutual inductance.) It is the persistent current property of cryotron loops that furnishes the memory necessary for sequential behavior. This is the basic principle upon which the following discussion is based.

If it is desired to change y from 1 to 0, it is necessary to introduce resistance into all paths of the left-hand branch of Fig. 2, while maintaining a superconducting path through the right-hand branch. Similarly, a change from y = 0 to y = 1 can be effected by making all paths in the right-hand branch resistive and maintaining at least one superconducting path in the left-hand branch. Thus, Box A in Fig. 2 should be resistive and Box B should be superconducting when the transition from y = 1 to y = 0 is to occur, and A should be superconducting and B resistive when the inverse transition is to take place.

The step from the flow table to the design of the circuit is made with the help of a transition matrix. The following paragraph describes the use of the transition matrix. This matrix is constructed from the excitation matrix by inserting 1 in the transition matrix for those entries of the excitation matrix in which the value of Y differs from the value of y, and 0 for those entries in which Y = y.

In other words, the entry is 1 in the transition matrix if, and only if, the secondary variable must change its value. The transition matrix corresponding to the excitation matrix of Fig. 1d is given in 3a.

It is a straightforward step from the transition matrix to the final circuit. The 0 row of the matrix in Fig. 3a indicates that y can change from 0 to 1 only for inputs $x_1, x_2 = 0, 1$; therefore, the right-hand branch of the circuit for secondary variable y must be resistive only when $\bar{x}_1 x_2 = 1$. Similarly, row 1 indicates that y can change from 1 to 0 only when $x_1, x_2 = 0, 0$ or $1, 0$; hence, the left-hand branch must be resistive only when $\bar{x}_2 = 1$. The final circuit, shown in Fig. 3b, follows immediately. The lower part of 3b is the output circuit, which is based upon the output matrix of Fig. 1c. A comparison of Fig. 1e and 3b illustrates that design by use of the transition matrix yields simpler circuitry.

In the circuit of Fig. 3b there is no combination of input variables for which both branches become resistive simultaneously. This is true in general for circuits constructed by means of the transition matrix. A resistive left branch implies that if the secondary variable is 1 it must change to 0, and a resistive right branch implies that if the secondary variable is 0 it must change

y \ x_1x_2	00	01	11	10
0	0	1	0	0
1	1	0	0	1

Figure 3a - Transition matrix
for excitation matrix of Fig. 1d.

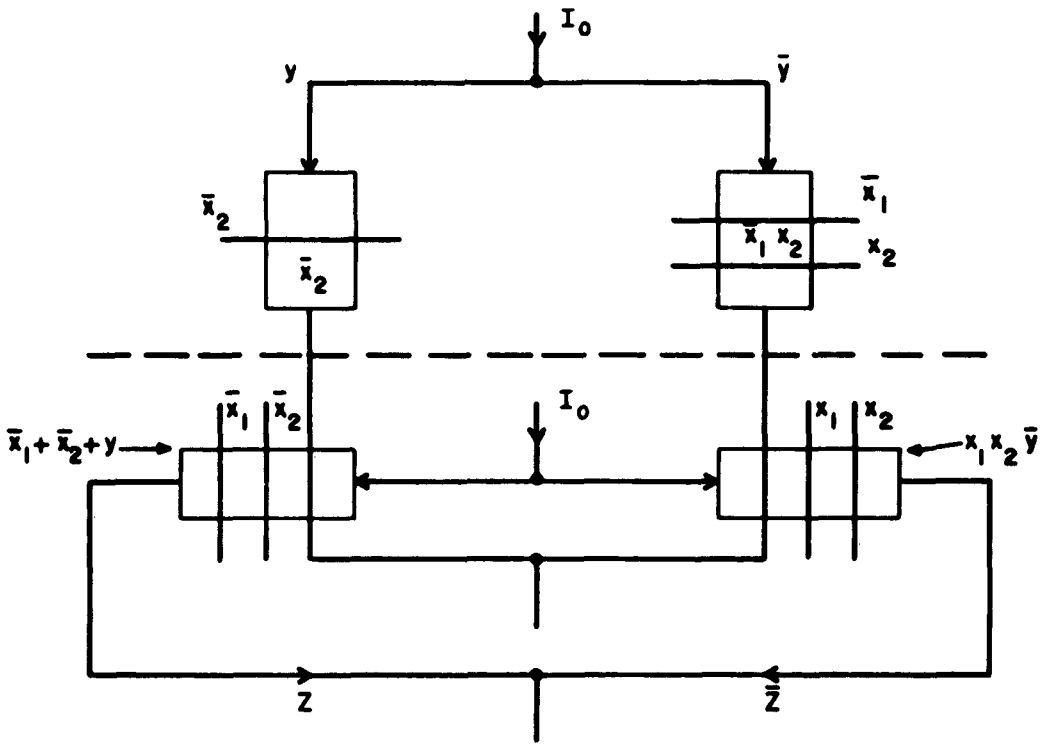


Figure 3b - Illustrative sequential circuit.

to 1. Hence, both branches being resistive implies that the secondary variable should switch from 0 to 1 to 0 to 1, and so on. In other words, the final circuit can be simultaneously resistive in both branches of a secondary variable circuit only if the original flow table contains a cycle. But, since a cryotron circuit will not cycle when both branches are resistive, true cycles can be built into cryotron circuits only when the cycles involve two or more secondary variables (pass through four or more different secondary states).

It will be instructive to consider a particular cryotron circuit for which the two inputs never equal 1 simultaneously. Output Z is to equal 1 after input W_1 is set equal to 1 and to remain 1 until input W_0 is set equal to 1, after which $Z = 0$ until $W_1 = 1$, and so on. This is a flip-flop, W_1 and W_0 being the set and reset pulses. The description is represented by the flow table of Fig. 4a, the merged flow table of 4b, the transition matrix of 4c, and the output matrix of 4e. Because the input combination $W_0, W_1 = 1, 1$ never occurs, the third column of the transition matrix may be completed in any convenient manner at all. The choice made for the third column entries is the one that yields the simplest algebraic expressions for the logical elements in the two

W_0W_1 : 00	01	11	10	OUTPUT Z
(1)	3	–	4	0
(2)	3	–	4	1
2	(3)	–	–	1
1	–	–	(4)	0

Figure 4a - Flow table.

W_0W_1 : 00	01	11	10
(1)	3	–	(4)
(2)	(3)	–	4

Figure 4b - Merged flow table.

y \ W_0W_1	00	01	11	10
0	0	1	–	0
1	0	0	–	1

Figure 4c - Transition matrix.

y \ W_0W_1	00	01	11	10
0	0	1	1	0
1	0	0	1	1

Figure 4d - Completed transition matrix.

y \ W_0W_1	00	01	11	10
0	0	0	0	0
1	1	1	1	1

Z

Figure 4e - Completed output matrix.

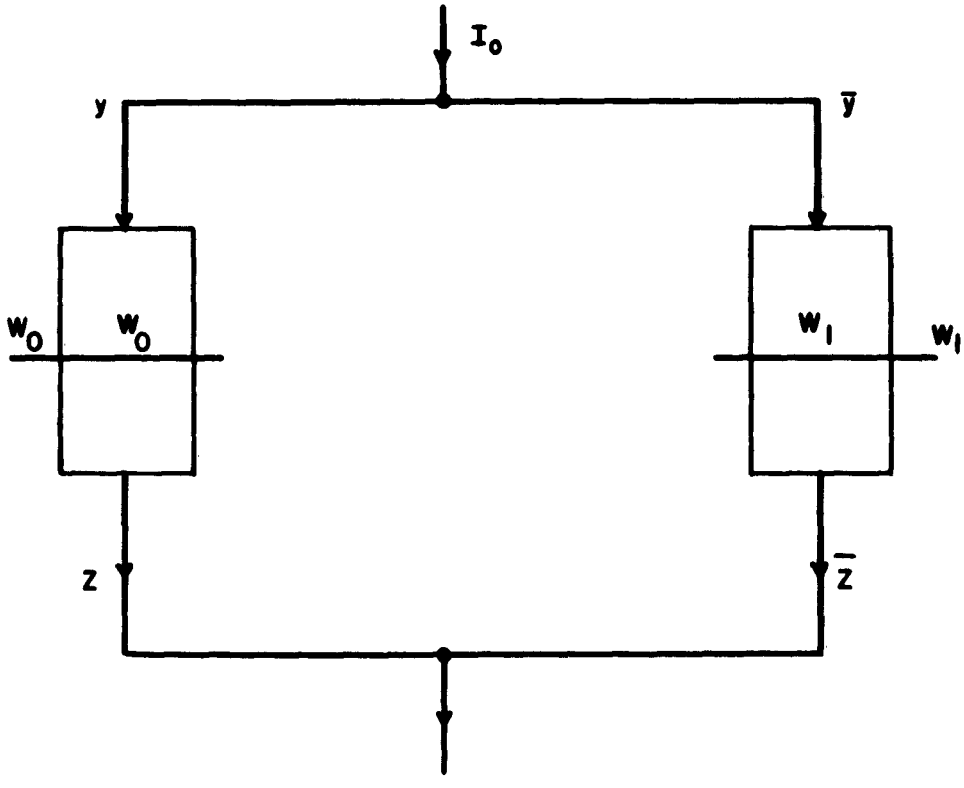


Figure 4f - Flip-flop.

X_1X_2: 00	01	11	10	OUTPUT Z
①	2	–	7	0
1	②	3	–	0
–	5	③	4	0
1	–	6	④	1
1	⑤	6	–	0
–	5	⑥	7	0
1	–	6	⑦	0

Figure 5a - Flow table.

X_1X_2: 00	01	11	10
①	②	3	7
–	5	③	4
1	–	6	④
1	⑤	⑥	⑦

Figure 5b - Merged flow table.

y_1 y_2 \ X_1X_2	00	01	11	10
00	0	0	0	0
01	–	0	0	–
11	–	–	–	1
10	0	0	0	0

Z

Figure 5c - Output matrix.

y_1 y_2 \ X_1X_2	00	01	11	10
00	00	00	01	10
01	00	11	01	11
11	00	10	10	11
10	00	10	10	10

Y_1 Y_2

Figure 5d - Excitation matrix.

	y_2 \ X_1X_2	00	01	11	10
$y_1=0$	0	0	0	0	1
	1	0	1	0	1
$y_1=1$	1	1	0	0	0
	0	1	0	0	0

Figure 5e - Transition matrix for y_1.

	y_1 \ X_1X_2	00	01	11	10
$y_2=0$	0	0	0	1	0
	1	0	0	0	0
$y_2=1$	0	1	0	0	0
	1	1	1	1	0

Figure 5f - Transition matrix for y_2.

branches of the y circuit. The completed transition matrix is given in Fig. 4d. The output matrix of Fig. 4e is also the result of making simplifying choices for optional entries. The final circuit is shown in Fig. 4f. This, of course, is a cryotron flip-flop.[6, 7, 8] Notice that both branches are resistive only when W_0, W_1 = 1, 1, which is an unallowed condition.

The preceding examples have concerned circuits with a single secondary variable. The next step in this exposition is the

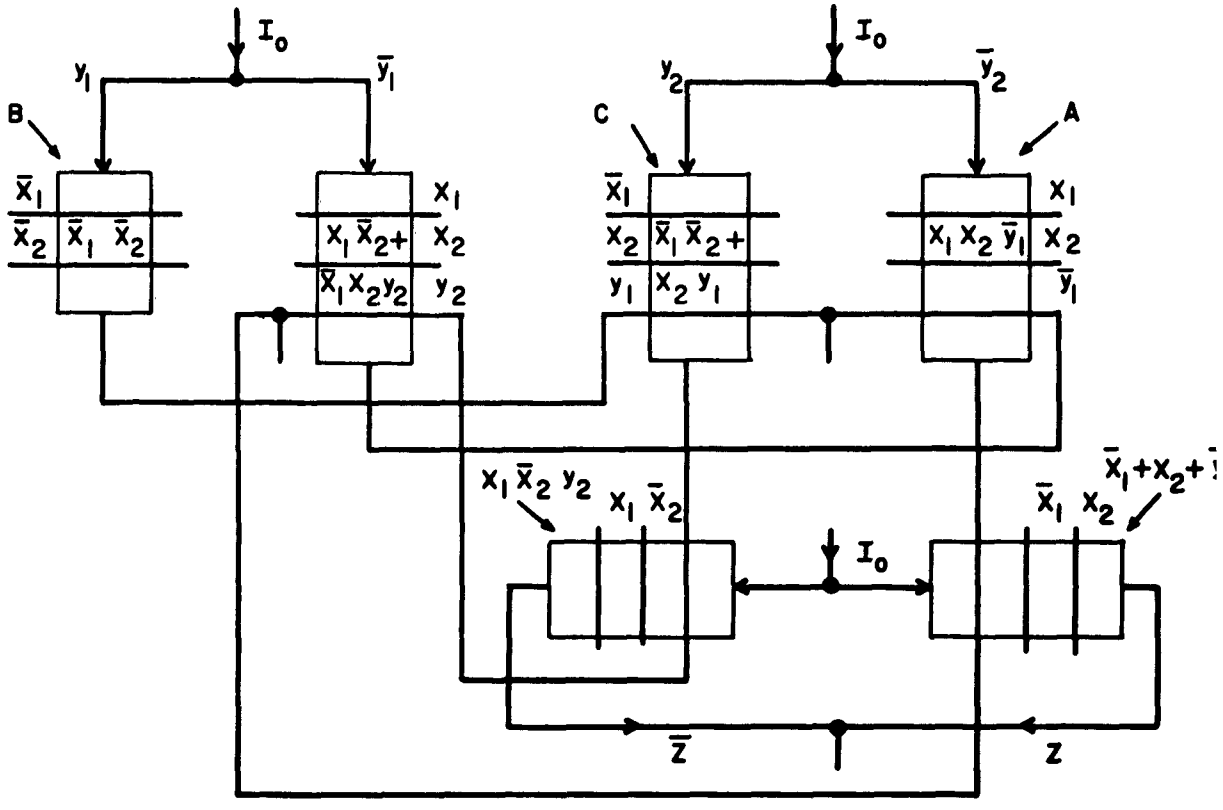


Figure 5g - Illustrative cryotron circuit.

demonstration of how sequential circuits with more than one secondary variable are constructed. Figure 5 refers to a circuit that yields an output whenever the input sequence 00-01-11-10 has occurred. The output is produced during the 10 phase of the sequence. Only one input variable may change at a time. Figure 5a is a flow table, 5b a merged flow table, 5c an output matrix, and 5d an excitation matrix. The excitation matrix allows a noncritical race in the 00 column, and provides for the secondary variable transition 01-11-10 in the 01

column. The corresponding transition matrices are given in Fig. 5e and 5f in a convenient form. The final circuit is easily constructed. To see how it is constructed, consider, for example, the upper half of Fig. 5f, the transition matrix for y_2. This submatrix pertains to the cases in which the present state of y_2 is 0. The single 1 indicates that y_2 is to change from 0 to 1 only when $x_1, x_2, y_1 = 1, 1, 0$. This statement corresponds to box A in Fig. 5g in the right-hand branch of the circuit for y_2. The submatrix in Fig. 5f for which $y_2 = 1$ has 1 entries where $\bar{x}_1\bar{x}_2 + x_2y_1 = 1$. These are the conditions for which y_2 must change from 1 to 0. Thus, the left-hand branch of the y_2 circuit in Fig. 5g is resistive when $\bar{x}_1\bar{x}_2 + x_2y_1 = 1$. In a similar manner, the circuit for y_1 is easily derived from Fig. 5e. A completed output matrix leads to the output circuit in Fig. 5g.

The procedure followed in order to arrive at Fig. 5g is completely general and applies to any sequential circuit with any number of secondary variables. Once the description of desired operation for a cryotron circuit is given, the excitation matrices can be constructed according to well-known principles.[2, 5] The transition matrices follow immediately. The circuits for the secondary variables are designed from the corresponding transition matrices. This is the crux of the design procedure. The transition matrix for

secondary variable y_i is divided into two submatrices. One submatrix corresponds to those rows of the y_i transition matrix for which $y_i = 0$. The rows of this submatrix are labeled with the state designations of the remaining secondary variables. The logical function represented by this submatrix is used to design a box for one of the two branches of the y_i circuit, the one that carries current representing $\bar{y}_i$. The box must be resistive only when the function equals 1. The other submatrix corresponds to the rows of the transition matrix for which $y_i = 1$. It is used to design the box for the other branch of the y_i circuit.

Before restrictions and hazards are discussed, three topics related to the synthesis procedure using the transition matrix will be mentioned briefly.

1. Any simple cryotron circuit of the type in Fig. 2 will perform in a sequential fashion unless the functions represented by blocks A and B are logical duals. If this fact is ignored by the designer of logical circuitry, then incorrect operation is likely to result.

2. The elementary cryotron circuit is closely related to the set-reset flip-flop. The essential difference is that the cryotron circuit is a dc element, whereas the set-reset flip-flop is a pulsed

device with a delay assumed to be greater than the pulse duration.

3. An example is used to illustrate an unexpected feature of the circuit discussed here. Figure 6 shows an output circuit that is simpler than, but equivalent to, the output circuit of Fig. 5g. Because a transition from stable state 4 (see Fig. 5b) does not occur without secondary variable y_2 becoming 0, the transition from Z = 1 to Z = 0 occurs when y_2 becomes 0. Hence, all that is necessary in the right-hand branch of the output circuit is a cryotron that is resistive when $y_2 = 0$. Similarly, the transition from Z = 0 to Z = 1 occurs only when $y_2 = 1$ and $x_2 = 0$; therefore, it is sufficient to include a box in the left-hand branch of the output circuit that is resistive when $\bar{x}_2 y_2 = 1$. This simplification procedure is quite novel. Unfortunately, the concept of making use of the memory property of the output circuit has not yet been systematized.

RESTRICTION ON LOGICAL DESIGN

There is one restriction on the way in which sequential cryotron switching circuits can be realized, as the following example illustrates. Consider box B in the left-hand branch of the y_1 circuit in Fig. 5g. Assume that it is realized by the configuration

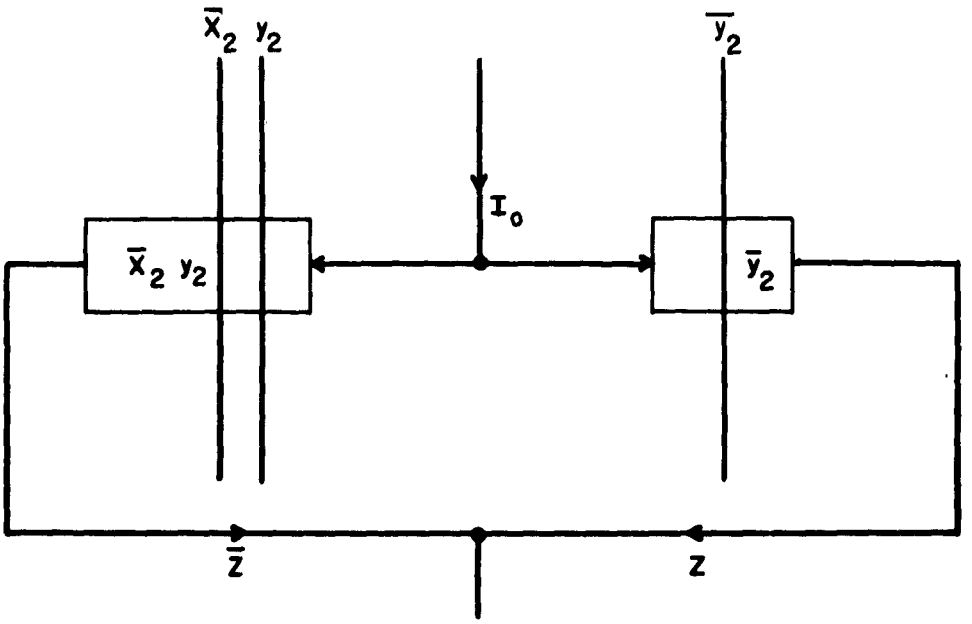


Figure 6 - Alternative to the output circuit of Fig. 5g.

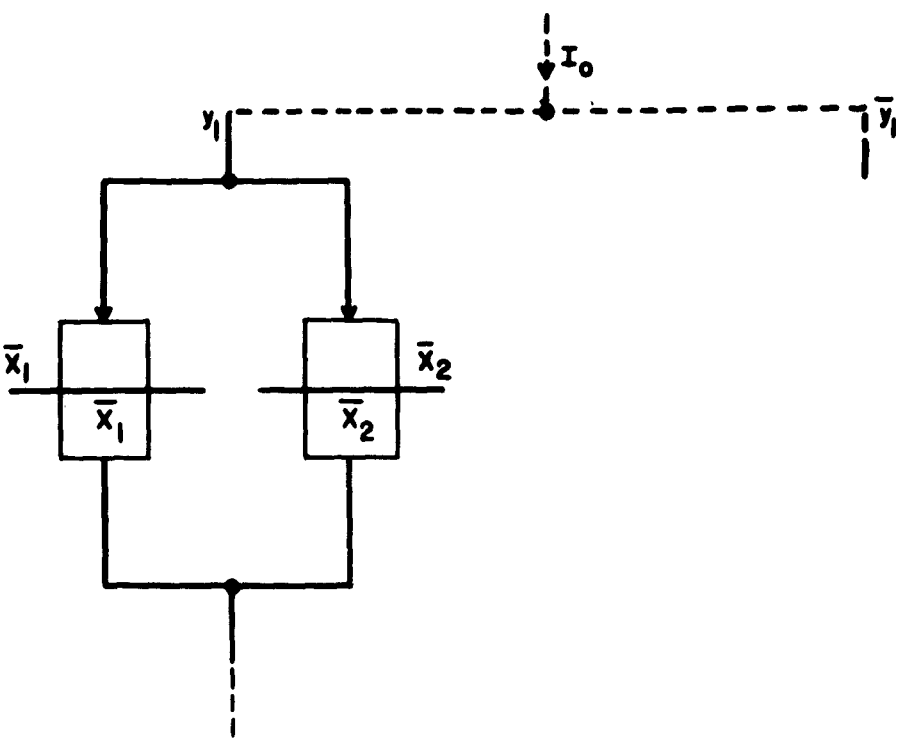


Figure 7- Illustration of an unallowed condition.

of Fig. 7. Let the secondary variables y_1, y_2 be equal to 1, 0, and assume that $x_1, x_2 = 1, 0$. Thus current I_o flows through cryotron $\bar{x}_1$, because cryotron $\bar{x}_2$ is resistive. When x_2 changes from 0 to 1 both cryotrons are superconductive, but current I_o continues to flow through cryotron $\bar{x}_1$. A difficulty arises when x_1 next changes from 1 to 0, since cryotron $\bar{x}_1$ becomes resistive. Current I_o is forced to change paths and divide between the $\bar{x}_2$ cryotron and the right-hand branch of the y_1 circuit. The way this division occurs depends on the relative inductances of the superconducting paths. An error can occur if the current should flow in the right-hand branch of the y_1 circuit. Such an error can always be prevented if no closed loops exist in any branch. This is equivalent to saying that the cryotron networks in both branches of all secondary variable circuits must be series connections of gates. This is a sufficient condition for the elimination of false operation. The reader who wonders how logical multiplication can be realized without parallel cryotrons is reminded of the existence of multiple-control cryotrons, which can perform both logical addition and multiplication.[3]

HAZARDS

The following discussion illustrates that circuit action may not go to completion if more than one secondary variable must change as a result of a change of input variable. This is a shortcoming of circuits designed via the transition matrix. For example, the excitation matrix of Fig. 5d demands that when the circuit is in state $y_1, y_2 = 0, 1$ and when x_1, x_2 changes from 1, 1 to 0, 1, then y_1, y_2 must change from 0, 1 to 1, 1 to 1, 0 before reaching a stable state.

Figure 8a is a convenient diagram for the purpose of studying this situation, for the figure is equivalent to the secondary variable portion of Fig. 5g when $x_2 = 1$. Figure 8b shows the currents i_1 and i_2 of Fig. 8a as a function of time. Initially, $x_1 = 1$, $i_1 = 0$, $i_2 = I_o$, cryotrons 1 and 2 are superconductive, and cryotron 3 is normal. When x_1 changes to 0 at time $t = 0$, cryotron 1 goes normal, 2 remains superconductive, and 3 switches superconductive. As a result, i_1 increases exponentially from 0 and approaches I_o. Current i_2 remains constant at I_o. When i_1 becomes large enough, cryotron 2 switches to the normal state at some time $t = t_1$. This causes i_2 to decrease exponentially from I_o and approach 0. When i_2 becomes

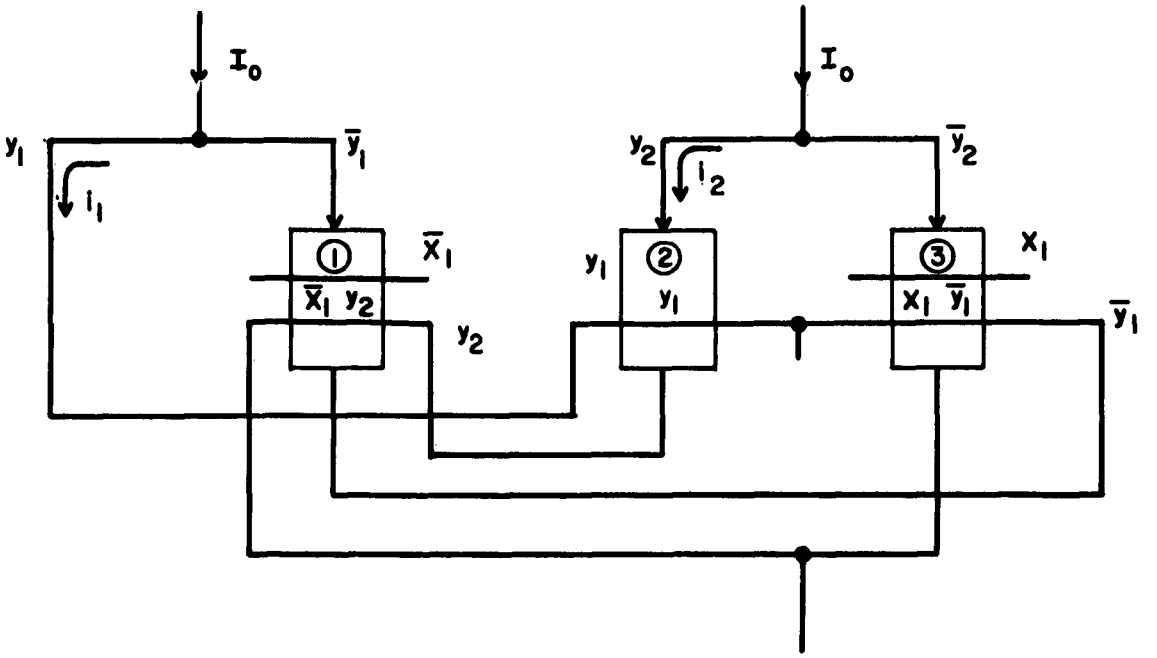


Figure 8a - Illustration of change of two secondary variables.

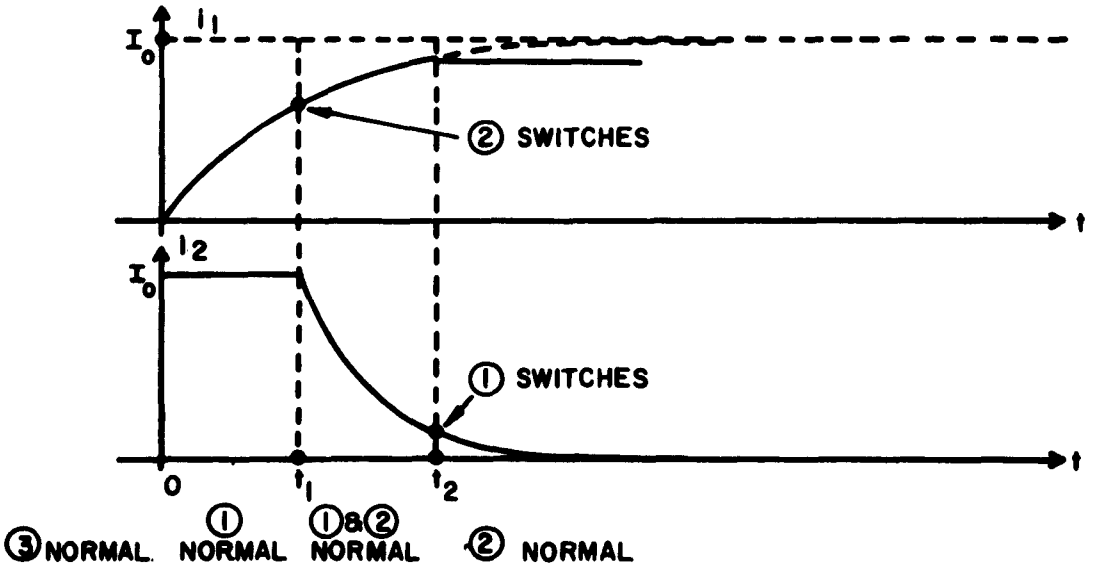


Figure 8b - Currents in Fig. 8a.

small enough so that cryotron 1 becomes superconductive again at some time $t = t_2$, current i_1 remains constant at the value it assumes at time t_2. Current i_1 does not complete the transition to I_o. In this particular circuit, there is no malfunction because of the incomplete switching. This fact can be concluded after careful analysis of Fig. 5g. This analysis consists of examining all the possible sequences of inputs after time t_2. In general, however, incorrect behavior may occur unless the time constants are arranged so that currents attain sufficiently large values before the cryotrons driving the currents positive become superconductive.

Another solution is to avoid multiple changes of secondary states. The latter can always be done by an appropriate selection of secondary state assignments,[9] although such a selection may greatly increase the size of the circuit. Often, however, multiple changes of secondary variables can be avoided if care is taken in selecting a merged flow table. The merged flow table of Fig. 9 can be implemented instead of that of Fig. 5b. The circuit resulting from Fig. 9 displays complete switching.

One more remark should be made about the circuit of Fig. 5g. A noncritical race occurs when the condition of inputs $x_1, x_2 = 0, 0$ causes the secondary variables to change from

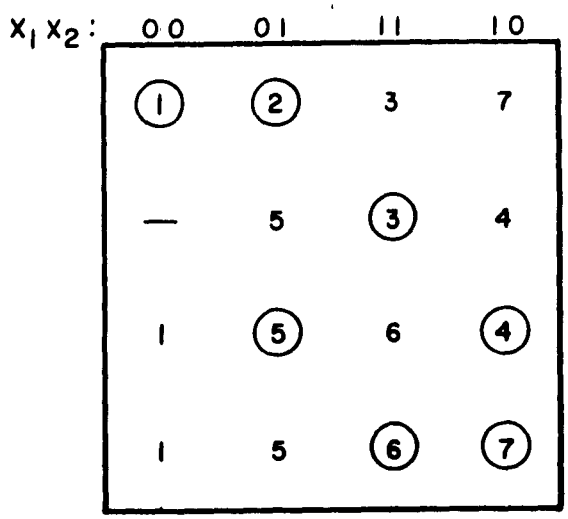


Figure 9 - Alternative merged flow table for Fig. 4.

y_1, y_2 = 1, 1 to 0, 0. The currents do go to completion in this case, although both secondary variables must change their values. This can be seen by inspecting the figure.

The conventional problems of hazardous conditions arise in sequential cryotron switching circuits, and the techniques for eliminating these hazards are the same as those used with switching circuits built of other components. Both static[10] and essential[11] hazards will be illustrated.

The reader may have noticed the "static tie-set hazard"[10] that can exist in box C in the left-hand branch of y_2 in the circuit of Fig. 5g. This is indicated by the disjunctive circled 1's in Fig. 5f.

When $x_1, y_1, y_2 = 0, 1, 1$, and x_2 changes from 0 to 1, the resistance in the left-hand branch of y_2 can change momentarily from R to 0 and then back to R if the time constants in the circuit are critically arranged. Fortunately, the tie-set hazard has no effect upon the circuit operation, since a temporarily superconducting path cannot unswitch current already flowing in the opposite branch. Nevertheless, a branch with a hazard in its cut-set, for which the resistance changes momentarily from 0 to R and back to 0, can cause incorrect operation. For example, if box C in the left-hand branch of y_2 in Fig. 5g is realized by the cryotron configuration of Fig. 10, then when $x_1, y_1, y_2 = 1, 0, 1$, and x_2 changes from 1 to 0, the branch becomes momentarily resistive if the x_2 cryotron becomes normal before the $\bar{x}_2$ cryotron goes superconductive. If this occurs, some current may be switched to the right-hand branch of the y_2 circuit. If enough current is switched, false operation can result. This cut-set hazard is indicated by the disjunctive circled 0's in Fig. 5f. This critical type of static hazard is eliminated if appropriate cut-sets are added to the circuit or if each branch is a series connection of cryotrons. (The circuit of Fig. 10, used to realize the left-hand branch of y_2 in Fig. 5g, is a parallel connection of combinations of cryotrons.)

Essential hazards[11] can also occur in sequential cryotron

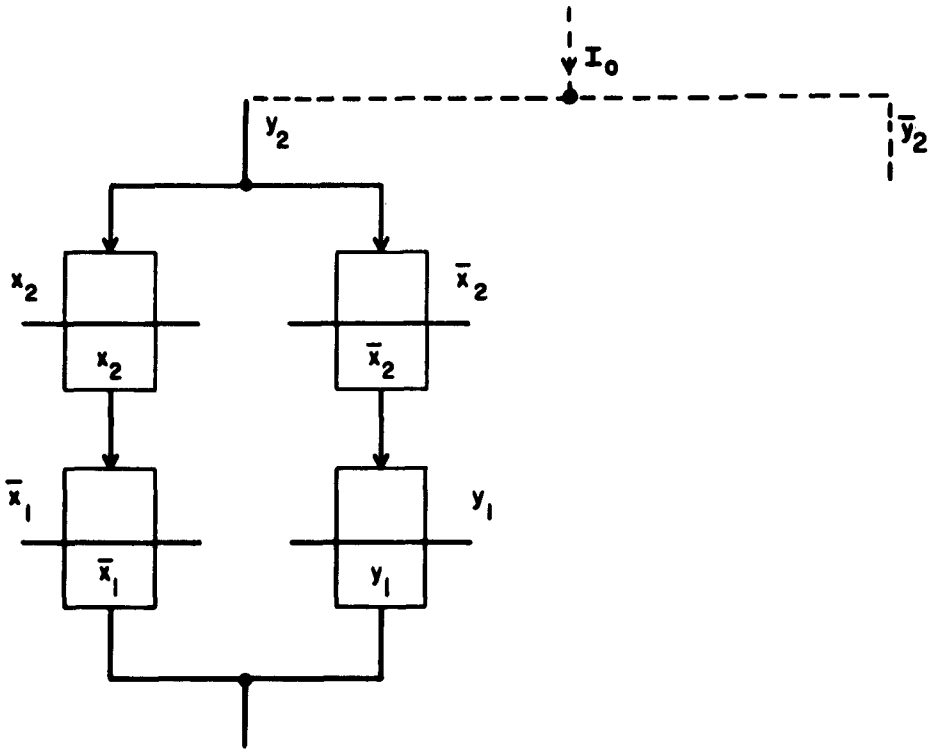


Figure 10 - Circuit in which a static cut-set hazard exists.

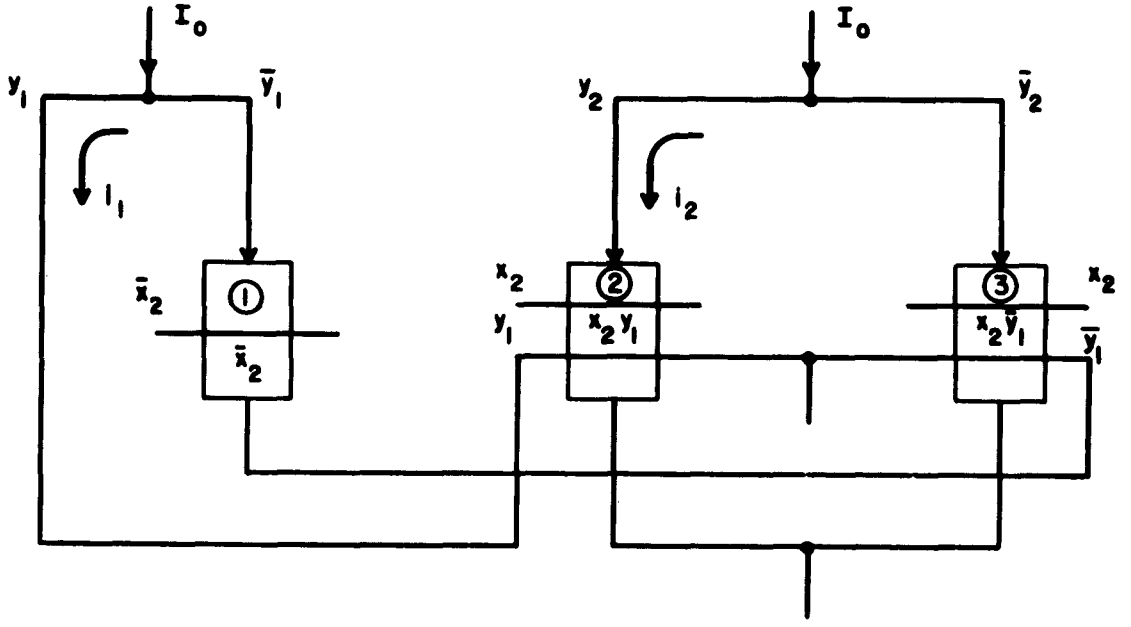


Figure 11 - Illustration of an essential hazard in the circuit of Fig. 5g.

switching circuits. This type of hazard, to be illustrated by an example, is especially bothersome in cryotron circuits. Since these circuits can have extremely fast switching speeds, the propagation time of signals can be an appreciable fraction of the switching time of a cryotron gate. This fact implies that significantly large unwanted delays can exist in cryotron circuits. The probability that essential hazards exist can be relatively large unless steps are taken to prevent their existence.

Consider, for an example of an essential hazard, the case in which the circuit of Fig. 5g is initially in state $y_1, y_2 = 0, 1$, and the inputs x_1, x_2 are changed from 1, 1 to 1, 0. Since x_1 remains constant at 1, the equivalent circuit of Fig. 11 may be used to study the situation. Initially, $i_1 = 0$, $i_2 = I_o$, cryotrons 1 and 2 are superconductive, and cryotron 3 is normal. When x_2 changes from 1 to 0, cryotron 1 should go normal and cryotron 3 superconductive simultaneously. Assume, however, that the current representing $\bar{x}_2$ in cryotron 1 increases before the current representing x_2 in cryotrons 2 and 3 dies out. As a result of the switching of cryotron 1 to the normal state, current i_1 will begin to increase exponentially, approaching I_o. If the time constant in the loop of i_1 is small, i_1 may increase rapidly enough to cause cryotron 2 to switch normal

and cryotron 3 to switch superconductive before the current x_2 becomes very small. This, in turn, causes i_2 to decrease exponentially (approach 0) until the x_2 current becomes small enough for cryotron 2 to return to the superconductive state. If the time constants are unfavorable enough, current i_2 can become very small before the circuit reaches a stable state. This stable state is essentially $y_1, y_2 = 1, 0$. Correct behavior would carry the circuit to the state $y_1, y_2 = 1, 1$.

Essential hazards can be eliminated only by assuring that certain time constants are relatively large or by inserting delays in appropriate positions. This is a fundamental conclusion of Unger's work. In the example the hazard can be removed by guaranteeing that the current representing x_2 becomes sufficiently small before the current representing $\bar{x}_2$ exceeds a critical value, or else by making the time constant of the loop of x_2 sufficiently smaller than that of either current loop i_1 or i_2.

CONCLUSION

A brief investigation of the relationship between synthesis via the transition matrix and synthesis by means of the direct design

procedure will indicate the significance of the former. Figure 12 is a redrawing of Fig. 1e so that certain features are made prominent. Figure 12 can be simplified on the basis of the following analysis. Consider, for example, the block labeled $x_1\bar{y}$ in the circuit in Fig. 12. This block is to be resistive when $x_1 = 1$ and $y = 0$; under these circumstances current will be shifted from the left branch of the y circuit to the right. But if current flows in the left branch, then $y = 1$, and the block $x_1\bar{y}$ cannot be resistive. As a result, that block is superfluous. Another way of viewing this question is to consider the $x_1\bar{y}$ cryotron as a latching device that prevents current from flowing in the left branch whenever $x_1 = 1$ and y should equal 0. The box is unnecessary, however, because of the persistent current property of the cryotron loop. Similar reasoning leads to the conclusion that the block labeled x_2y in the right branch is also superfluous. Elimination of these two blocks leads to the circuit of Fig. 3b.

In general, the simplification procedure just illustrated, when applied to circuits designed by the direct procedure, will lead to circuits very similar to those designed by the transition matrix technique. The main advantage of design by use of the transition matrix, aside from the simplicity of the approach, is more efficient circuit simplification when "never occur" input conditions must be

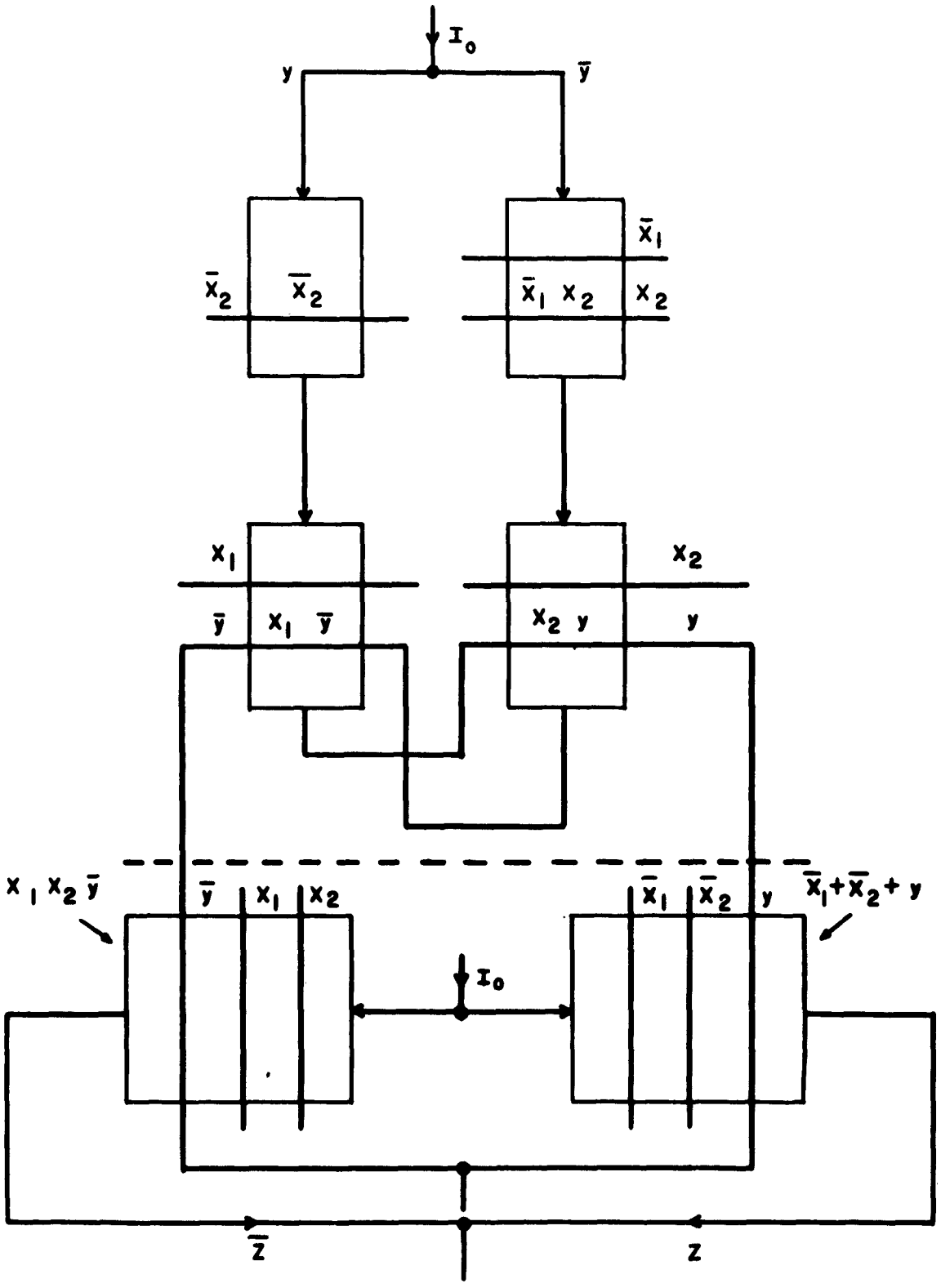


Figure 12 - Circuit of Fig. 1e redrawn.

handled. Furthermore, design via the transition matrix calls attention to the deliberate use of the memory found in cryotron loops to design sequential cryotron switching circuits.

REFERENCES

1. D. A. Buck, "The Cryotron - A Superconductive Computer Component," Proc. Inst. Radio Engrs., Vol. 44, No. 4, April 1956, pp. 482-493.

2. S. H. Caldwell, "Switching Circuits and Logical Design," (John Wiley and Sons, Inc., New York, 1958).

3. A. L. Leiner, "Cryotron High-Speed Memory Systems," Project Lightning, Eighth Quarterly Progress Report, November 1960, pp. 70-78.

4. M. K. Haynes, "Cryotron Storage Arithmetic and Logical Circuits," Project Lightning, Sixth Quarterly Progress Report, Vol. III, May 1960, pp. 58-78.

5. D. A. Huffman, "The Synthesis of Sequential Switching Circuits," J. Franklin Inst., Vol. 257, No. 3, March 1954, pp. 161-190; Vol. 257, No. 4, April 1954, pp. 275-303.

6. A. E. Slade, "Thin-Film Cryotrons; Part II - Cryotron Characteristics and Circuit Applications," Proc. Inst. Radio Engrs., Vol. 48, No. 9, September 1960, p. 1572.

7. N. Rochester, "Fast Memory Using Simple Loops with In-Line Cryotrons," IBM Cryogenic Program, Second Supplemental Progress Report, November 1960, p. 165.

8. D. R. Young, "Superconducting Circuits," Research Report RC-54, IBM Research Center, Poughkeepsie, New York, May 1958, p. 53.

9. D. A. Huffman, "A Study of the Memory Requirements of Sequential Switching Circuits," Technical Report 293, Research Laboratory of Electronics, MIT, March 1955.

10. D. A. Huffman, "The Design and Use of Hazard-Free Switching Networks," Assoc. for Computing Machinery, Vol. 4, No. 1, January 1957, pp. 47-62.

11. S. H. Unger, "Hazards and Delays in Asynchronous Sequential Switching Circuits," IRE Trans. on Circuit Theory, Vol. CT-6, No. 1, March 1959, pp. 12-25.

APPENDIX VIII

General Lumped-Constant Analysis of Three-Branch Cryogenic Loops

E. D. Conroy

ABSTRACT

When a multi-branched cryogenic loop switches, transient currents flow in all branches. This appendix derives general expressions for these currents in a three-branch loop in terms of branch parameters and examines the special case in which all branch inductances and resistances are identical. It is shown that under certain conditions the presence of transient back current can improve both the speed and power characteristics of the loop.

INTRODUCTION

When a multi-branched cryogenic loop switches, the current leaving the branch becoming resistive will temporarily divide among all other branches of the loop. Ultimately, current will flow only in the superconducting branches, but in the interim, transient back currents will flow in the other resistive branches. The nature of these back currents is important, for in signal branches they can be interpreted as false outputs if they become too large. This report derives general expressions for the branch currents in a three-branch loop, and discusses some of the effects of the back current.

The analysis assumes that no voltage due to mutual inductance is induced during switching. It also assumes that the inductance and resistance in the loop may be treated as lumped parameters. If the parameters are such that the branch resistances are matched to the characteristic impedance of the films defining the loop, transmission-line effects must be considered.

BRANCH CURRENTS

The loops considered will have three parallel branches, but each branch will be a series connection of gates; i. e., no branch will contain sub-loops with the possibility for persistent circulating currents.

Such a loop is shown at the beginning of a switch in Fig. 1. The first branch is initially carrying the entire supply current, I_s, and this branch becomes resistive at time t = 0. At this time the second branch is superconducting and the third branch resistive. The branch parameters are the equivalent series inductance and resistance of their respective branches.

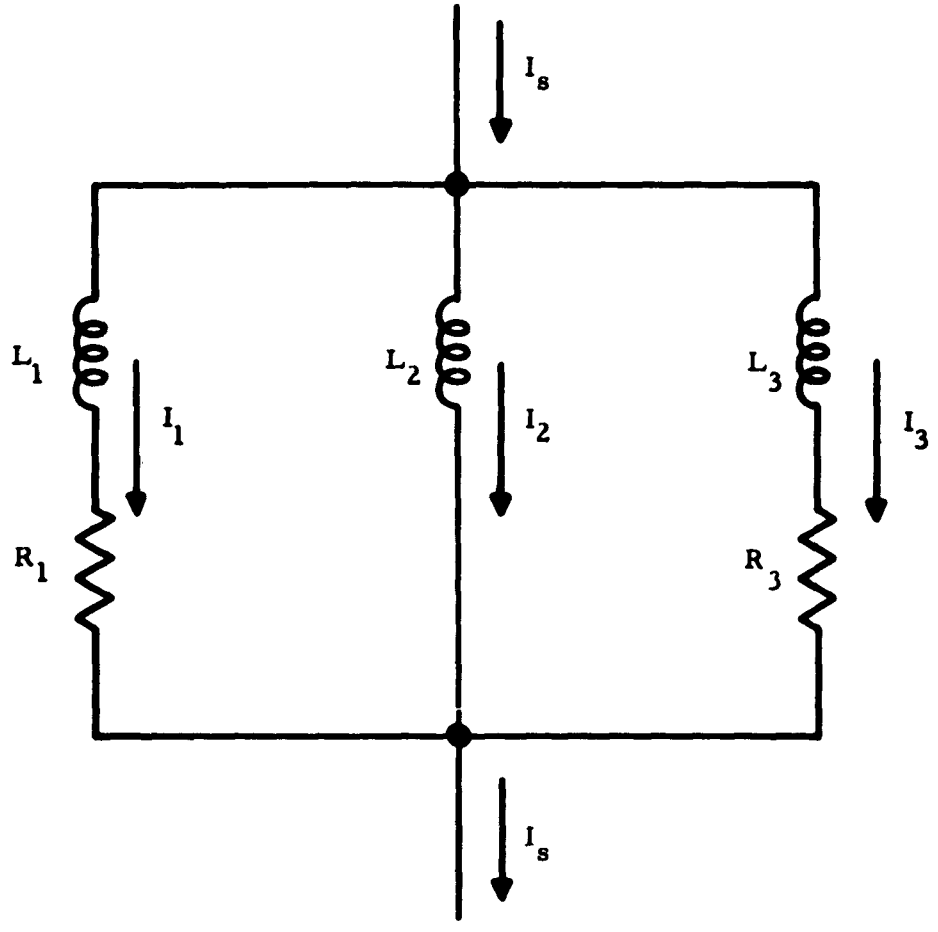


Figure 1 - Three-branch cryogenic loop.

The equations for the loop of Fig. 1 are

$$L_1 \frac{di_1}{dt} + R_1 i_1 = L_2 \frac{di_2}{dt}, \tag{1}$$

$$L_3 \frac{di_3}{dt} + R_3 i_3 = L_2 \frac{di_2}{dt}, \tag{2}$$

and

$$i_1 + i_2 + i_3 = I_s. \tag{3}$$

Applying the LaPlace transformation,

$$sL_1 I_1(s) - sL_2 L_2(s) + R_1 I_1(s) = L_1 I_s, \tag{4}$$

$$sL_3 I_3(s) - sL_2 I_2(s) + R_3 I_3(s) = 0, \tag{5}$$

$$sI_2(s) = I_s - sI_1(s) - sI_3(s). \tag{6}$$

Substituting (6) in (4) and (5),

$$s(L_1+L_2)I_1(s) + sL_2 I_3(s) + R_1 I_1(s) = (L_1+L_2)I_s, \tag{7}$$

$$s(L_2+L_3)I_3(s) + sL_2 I_1(s) + R_3 I_3(s) = L_2 I_s. \tag{8}$$

Solving (7) and (8) for $I_1(s)$ and $I_3(s)$ respectively,

$$I_1(s) = \frac{(L_1+L_2)I_s - sL_2I_3(s)}{s(L_1+L_2) + R_1}, \tag{9}$$

$$I_3(s) = \frac{L_2I_s - sL_2I_1(s)}{s(L_2+L_3) + R_3}. \tag{10}$$

Substituting (10) in (7) and multiplying both sides by $s(L_2+L_3)+R_3$,

$$I_1(s)[\,s^2(L_1L_2+L_1L_3+L_2L_3)+s(\{L_1+L_2\}R_3+\{L_2+L_3\}R_1) + R_1R_3]$$

$$= [\,s(L_1L_2+L_1L_3+L_2L_3) + (L_1+L_2)R_3]\,I_s\,. \tag{11}$$

Substituting (9) in (8) and multiplying both sides by $s(L_1+L_2)+R_1$,

$$I_3(s)[\,s^2(L_1L_2+L_1L_3+L_2L_3) +s(\{L_1+L_2\}R_3+\{L_2+L_3\}R_1) + R_1R_3]$$

$$= L_2R_1I_s\,. \tag{12}$$

If

$$\alpha \equiv \frac{(L_1+L_2)R_3 + (L_2+L_3)R_1}{(L_1L_2+L_1L_3+L_2L_3)}, \tag{13}$$

and

$$\beta^2 \equiv \frac{R_1R_3}{(L_1L_2+L_1L_3+L_2L_3)}, \tag{14}$$

then

$$I_1(s) = \frac{sI_s + \dfrac{(L_1+L_2)R_3 I_s}{(L_1L_2+L_1L_3+L_2L_3)}}{s^2 + \alpha s + \beta^2},$$

and

$$I_3(s) = \frac{\dfrac{L_2R_1 I_s}{(L_1L_2+L_1L_3+L_2L_3)}}{s^2 + \alpha s + \beta^2},$$

which can be written as

$$I_1(s) = I_s \frac{s}{\left(s+\frac{\alpha}{2}-\sqrt{\left(\frac{\alpha}{2}\right)^2-\beta^2}\right)\left(s+\left(\frac{\alpha}{2}\right)+\sqrt{\left(\frac{\alpha}{2}\right)^2-\beta^2}\right)} +$$

$$\frac{\dfrac{(L_1+L_2)R_3}{(L_1L_2+L_1L_3+L_2L_3)}}{\left(s+\frac{\alpha}{2}-\sqrt{\left(\frac{\alpha}{2}\right)^2-\beta^2}\right)\left(s+\frac{\alpha}{2}+\sqrt{\left(\frac{\alpha}{2}\right)^2-\beta^2}\right)} \qquad (15)$$

and

$$I_3(s) = \frac{\dfrac{L_2R_1 I_s}{(L_1L_2+L_1L_3+L_2L_3)}}{\left(s+\frac{\alpha}{2}-\sqrt{\left(\frac{\alpha}{2}\right)^2-\beta^2}\right)\left(s+\frac{\alpha}{2}+\sqrt{\left(\frac{\alpha}{2}\right)^2-\beta^2}\right)}. \qquad (16)$$

Making the inverse transformation,

$$i_1(t) = \frac{1}{2}\left[e^{-(\frac{a}{2}-\sqrt{(\frac{a}{2})^2-\beta^2})t} + e^{-(\frac{a}{2}+\sqrt{(\frac{a}{2})^2-\beta^2})t}\right] I_s + \left[\frac{(L_1+L_2)R_3-(L_2+L_3)R_1}{4\sqrt{(\frac{a}{2})^2-\beta^2}\ (L_1L_2+L_1L_3+L_2L_3)}\right]\cdot\left[e^{-(\frac{a}{2}-\sqrt{(\frac{a}{2})^2-\beta^2})t} - e^{-(\frac{a}{2}+\sqrt{(\frac{a}{2})^2-\beta^2})t}\right] I_s$$

(17)

$$i_3(t) = \left[\frac{L_2R_1}{2\sqrt{(\frac{a}{2})^2-\beta^2}\ (L_1L_2+L_1L_3+L_2L_3)}\right]\left[e^{-(\frac{a}{2}-\sqrt{(\frac{a}{2})^2-\beta^2})t} - e^{-(\frac{a}{2}+\sqrt{(\frac{a}{2})^2-\beta^2})t}\right] I_s.$$

(18)

Substituting (17) and (18) in (3) and solving for i_2,

$$i_2(t) = I_s - \frac{1}{2}\left[e^{-(\frac{a}{2}-\sqrt{(\frac{a}{2})^2-\beta^2})t} + e^{-(\frac{a}{2}+\sqrt{(\frac{a}{2})^2-\beta^2})t}\right]\left[\frac{(L_1+L_2)R_3+(L_2-L_3)R_1}{4\sqrt{(\frac{a}{2})^2-\beta^2}\ (L_1L_2+L_1L_3+L_2L_3)}\right]\left[e^{-(\frac{a}{2}-\sqrt{(\frac{a}{2})^2-\beta^2})t} - e^{-(\frac{a}{2}-\sqrt{(\frac{a}{2})^2-\beta^2})t}\right] I_s.$$

(19)

Equations (17), (18), and (19) give the branch currents as functions of the equivalent parameters of the branches.

BALANCED LOOPS

The branch current will now be evaluated for the special case of a three-branch loop with equal branch impedances in which current is switching from a resistive to a superconducting branch in the presence of a third resistive branch which initially carries no current.

$$L \equiv L_1 = L_2 = L_3. \tag{20}$$

$$R \equiv R_1 = R_3. \tag{21}$$

Substituting (20) and (21) into (13) and (14),

$$\alpha = \frac{4R}{3L}, \tag{22}$$

and

$$\beta^2 = \frac{R^2}{3L^2}. \tag{23}$$

The three branch currents are given by (17), (18), and (19). Substituting (20), (21), (22), and (23) into (17, (18), and (19),

$$i_1 = 0.5 \left(e^{-\frac{R}{3L}t} + e^{-\frac{R}{L}t} \right) I_s , \tag{24}$$

$$i_2 = \left(1 - e^{-\frac{R}{3L}t}\right) I_s , \tag{25}$$

and

$$i_3 = 0.5 \left(e^{-\frac{R}{3L}t} - e^{-\frac{R}{L}t} \right) I_s . \tag{26}$$

These currents are plotted in Fig. 2. The maximum back current occurs about one and one half time constants after switching and its value is less than 20 percent of the supply current, I_s. Thus, if the cryotrons are symmetrically biased (a reasonable assumption in view of the present variation in transition widths), back current could never cause a false output unless the effective transition width were greater than 60 percent of the signal swing. Furthermore, even if the transition region width is more than 60 percent of the signal swing, false outputs do not inevitably occur, but further analysis involving the time duration of the back current is required to determine the exact effect.

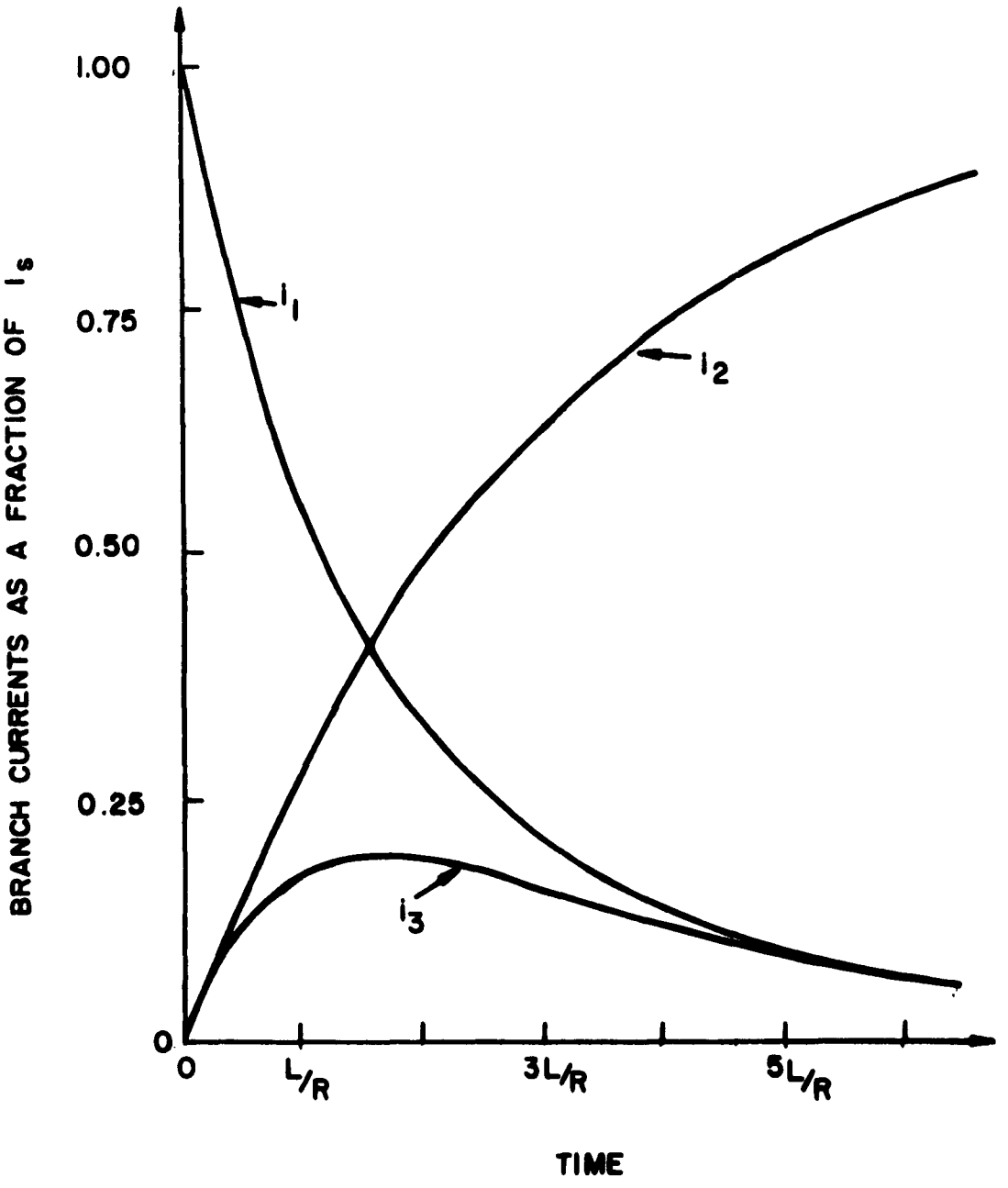


Figure 2 - Branch currents during switching of a three-branch loop.

SPEED

Figure 3 compares the signal-branch currents i_1 and i_2 from (24) and (25) with their values in a two-branch loop. The presence of the third branch is seen to speed the fall of the falling current and slow the rise of the rising current. This suggests the use of a resistive shunt, e.g., aluminum, across the loop as shown in Fig. 4. If the nature of the logic is such that the output is sensed by the absence of current in a branch, e.g., in Fig. 4 A. B is sensed by no current in the top branch and $\overline{A \cdot B}$ is sensed by no current in the middle branch, the effect of the back current in the shunt is to increase the speed of the circuit by causing the outputs to turn on more quickly.

Certain logical circuits, e.g., the OR tree in a memory, are used in such a manner that they require a fast switch in one direction only. They must recognize a certain condition with a minimum of delay, but having done so there is no harm in their having a relatively long recovery time before the next recognition. Under these circumstances a shunt can be used advantageously even if the output must be sensed in both branches, i.e., by the presence of current for some outputs and by the absence of current for others. The shunt is placed across the defining gate (as opposed to output gates) of the branch where current represents the time-critical signals.

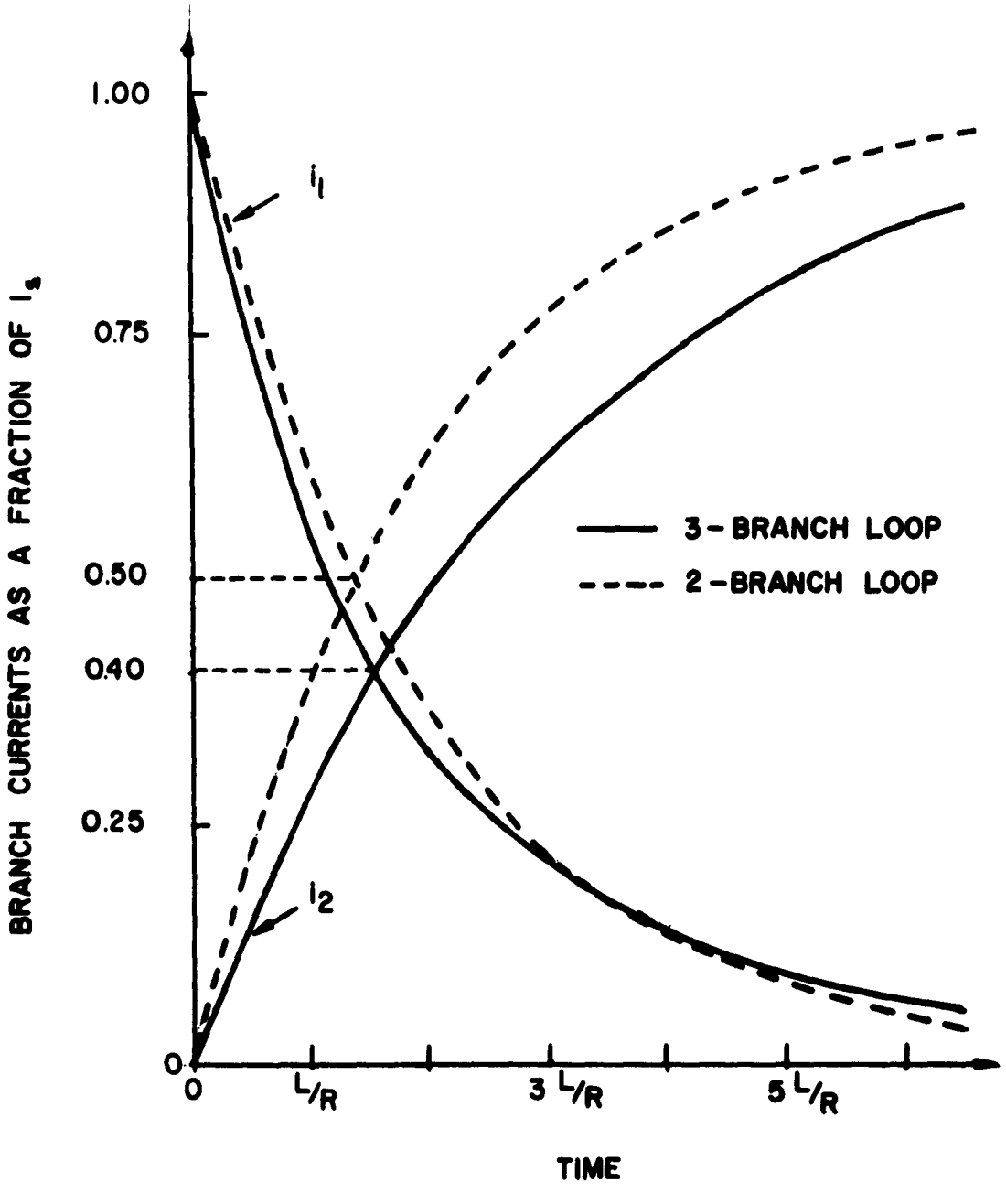


Figure 3 - Effect of third branch during switching (balanced branch impedance).

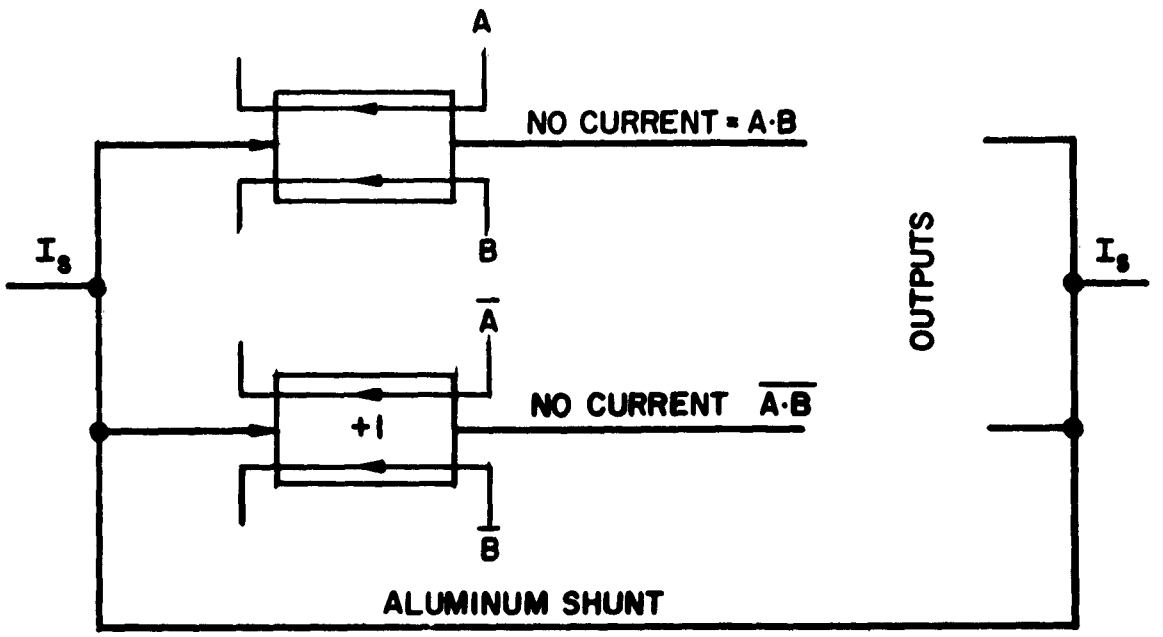


Figure 4 - Shunted cryogenic loop.

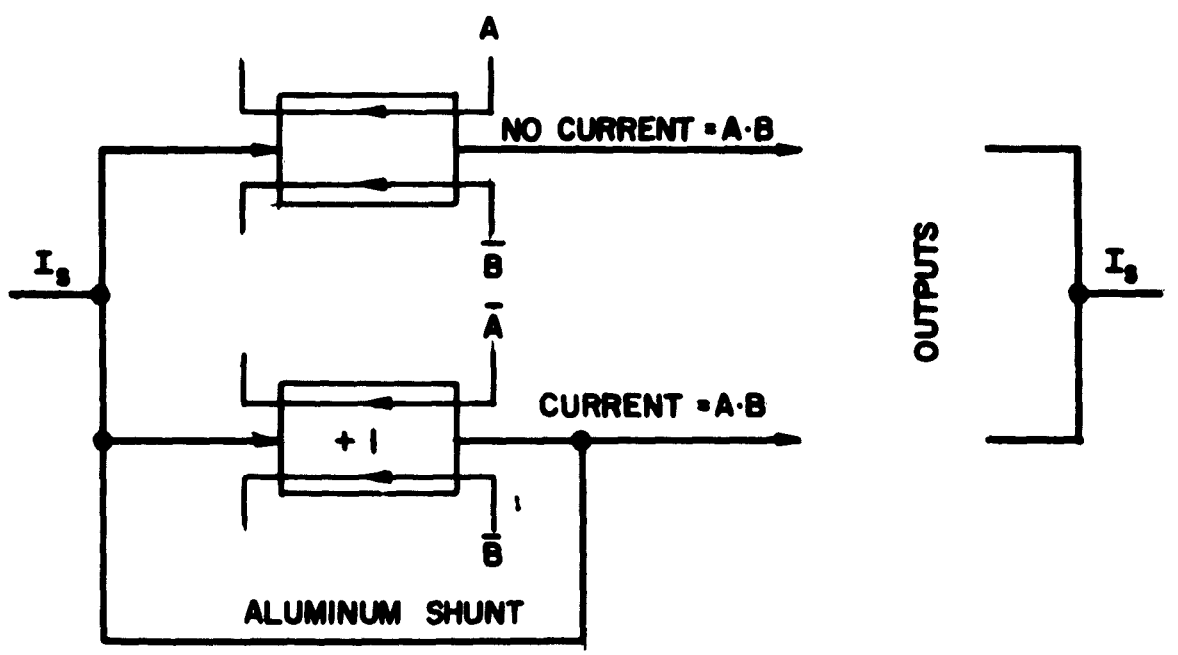


Figure 5 - Partially shunted cryogenic loop.

Figure 5 shows this configuration for a loop which recognizes $A \cdot B$, sensed in either branch, more quickly than $\overline{A \cdot B}$. Operation is similar to the loop shown in Fig. 4, although the effect is not as great as when the entire loop is shunted. While the shunting of a defining gate creates a subloop, in violation of one of the original assumptions, the fact that the aluminum is always resistive prevents any persistent circulating currents.

POWER

The presence of back current does not directly affect the total amount of energy dissipated during switching of a loop, although it does distribute the dissipation among the various resistive elements. From the conservation of energy

$$E_d = E_o + E_s - E_\infty, \tag{27}$$

where

E_d = total energy dissipated in the loop during switching,

E_o = energy stored in the loop at time $t=0$,

E_s = energy supplied by the power supply during the switch,

E_∞ = energy stored in the loop at time $t=\infty$.

With reference to Fig. 2, the energy stored in the loop at time $t=0$ is

$$E_o = \frac{1}{2} L_1 I_s^2. \tag{28}$$

The energy stored in the loop at time $t=\infty$ is

$$E = \frac{1}{2} L_2 I_s^2. \tag{29}$$

The power drawn from the power supply is

$$P_s = I_s V(t), \tag{30}$$

where $V(t)$ is the voltage across the loop. The energy supplied by the power supply during the switch is

$$E_s = \int_0^\infty I_s V(t) dt. \tag{31}$$

The voltage across the loop equals the voltage across any of the parallel branches. In terms of branch 2,

$$V(t) = L_2 \frac{di_2}{dt}. \tag{32}$$

Substituting (32) in (31),

$$E_s = L_2 I_s \int_0^\infty di_2,$$

or

$$E_s = L_2 I_s^2. \tag{33}$$

Equation (33) states that the energy supplied by the power supply during switching is independent of all loop parameters except the equivalent inductance of the superconducting branch to which the current is switching. Substituting (28), (29), and (33) into (27),

$$E_d = \frac{1}{2}(L_1+L_2) I_s^2. \tag{34}$$

Equation (34) states that the total energy dissipation is not a function of the parameters in branch 3 and thus equals the energy dissipation during switching of a two-branch loop (R_3 infinite).

Although the presence of back currents does not change the total energy dissipation, the back current, I_3, flowing through R_3 will obviously dissipate some amount of energy. The result is that less energy is dissipated in R_1. From the standpoint of heating, this means that the temperature rise of the gate going resistive will be less than in the two-branch case; some of the heat will be transferred to the bath and substrate by the back-current branch.

The preceding discussion pertains to the direct effects of back current on energy dissipation. There is also an indirect effect. In the above calculations it was assumed that one branch was always superconducting, i.e., the branch going superconductive switched before the branch going resistive switched. If this is not true and all branches are temporarily resistive, additional energy is drawn from the power supply. In a two-branch circuit with reasonable tolerances on the switching points it is quite possible to have all branches simultaneously resistive. Figure 3 shows that the presence of back current tends to speed the fall of current in the previous output line

and slow the rise of current in the new output line. For normal biasing, i.e., cryotrons biased superconducting, this has the effect of helping insure "make before break." (The inverse is true if the cryotrons are biased resistive.) With reference to Fig. 3, if cryotron gates are symmetrically biased, the two-branch loop will dissipate more energy than $\frac{1}{2}(L_1+L_2)I_s^2$ unless all cryotrons have identical switching points of exactly $0.5I_s$, a practical impossibility. For a normally biased three-branch loop this requirement is relaxed to $0.5I_s \pm 20$ percent. (For resistively biased cryotrons the effect is to guarantee additional dissipation unless these tolerances are exceeded.)

The action of the back current in insuring a "make before break" and distributing the energy dissipation suggests the use of the previously mentioned nonlogical resistive shunt to reduce gate temperature rise during switching, as well as for the possible speed advantage already discussed.

CONCLUSIONS

The presence of transient back current in a three-branch loop is to speed the fall of the falling current and slow the rise of the rising current. If certain constraints can be followed during logical design, therefore, back current can be used to increase the effective circuit speed. The back current also has two effects on the loop power dissipation. Although the total amount of energy dissipated during a switch is not changed, the dissipation is distributed between the falling-current branch and back-current branch. This reduces the heating in the resistive gate which is driving the supply current out of its original path. In addition, for normally biased cryotrons the offset between the falling and rising current helps to insure that no additional energy is dissipated because of simultaneous resistance in all branches.

REFERENCES

1. A. V. Shortell Jr., "Cryotron Multiposition Switches," M. S. Thesis in Electrical Engineering, MIT, May 1956.

2. D. A. Buck, "Superconductive Circuit Response Time: A Survey," Digest of Technical Papers, 1959 Solid-State Circuits Conference.

APPENDIX IX

A Control System for Testing an Experimental 16-Bit Memory Using In-Line Cryotrons

W. A. Notz, J. L. Smith, A. Weinberger

TABLE OF CONTENTS

FIGURES

TABLES

ABSTRACT

This paper describes a control system for testing an experimental high-speed cryotron memory. The control system utilizes 440 in-line cryotrons to provide an address generator, a data generator, system synchronization, monitoring facilities, and manual control. The logic of the system and the detailed cryotron circuits are specified in a series of diagrams. The operating speed of the system is summarized in a series of tables and diagrams. The paper concludes with the derivation of equations for the optimum fan-out of in-line cryotron circuits and some tables for the optimum fan-out of certain typical circuits.

INTRODUCTION

A control system for testing an in-line cryotron memory[1] was designed to provide the addresses, the information to be stored in the memory, the reading and writing signals for the memory, a memory-checking facility, and external system control and sensing facilities. The initial form of the memory was limited to 16 one-bit words so that the number of cryotrons would be kept small. The number of cryotrons in the control is also intended to be kept small. With this in mind, the control was designed so that its parts could be tested independently of one another and so that certain parts could be tested in minimum configurations with the memory. The external control provides the means by which an operator can start or stop the system or its parts, load the various storage cells of the control and vary the reading and writing rate. The sensing control provides the means of observing the operation of the system so that errors which might be internally detected can be traced to their source.

Figure 1 is a block diagram of the memory and its control. The functional subunits of the control are a clock (or stepping) pulse generator, an address generator, a read or write control circuit, an input shift-register, a serial unit adder, an output comparator, a full-cycle detector, and external controls and observation terminals. The system is intended to be operated one step at a time under push-button

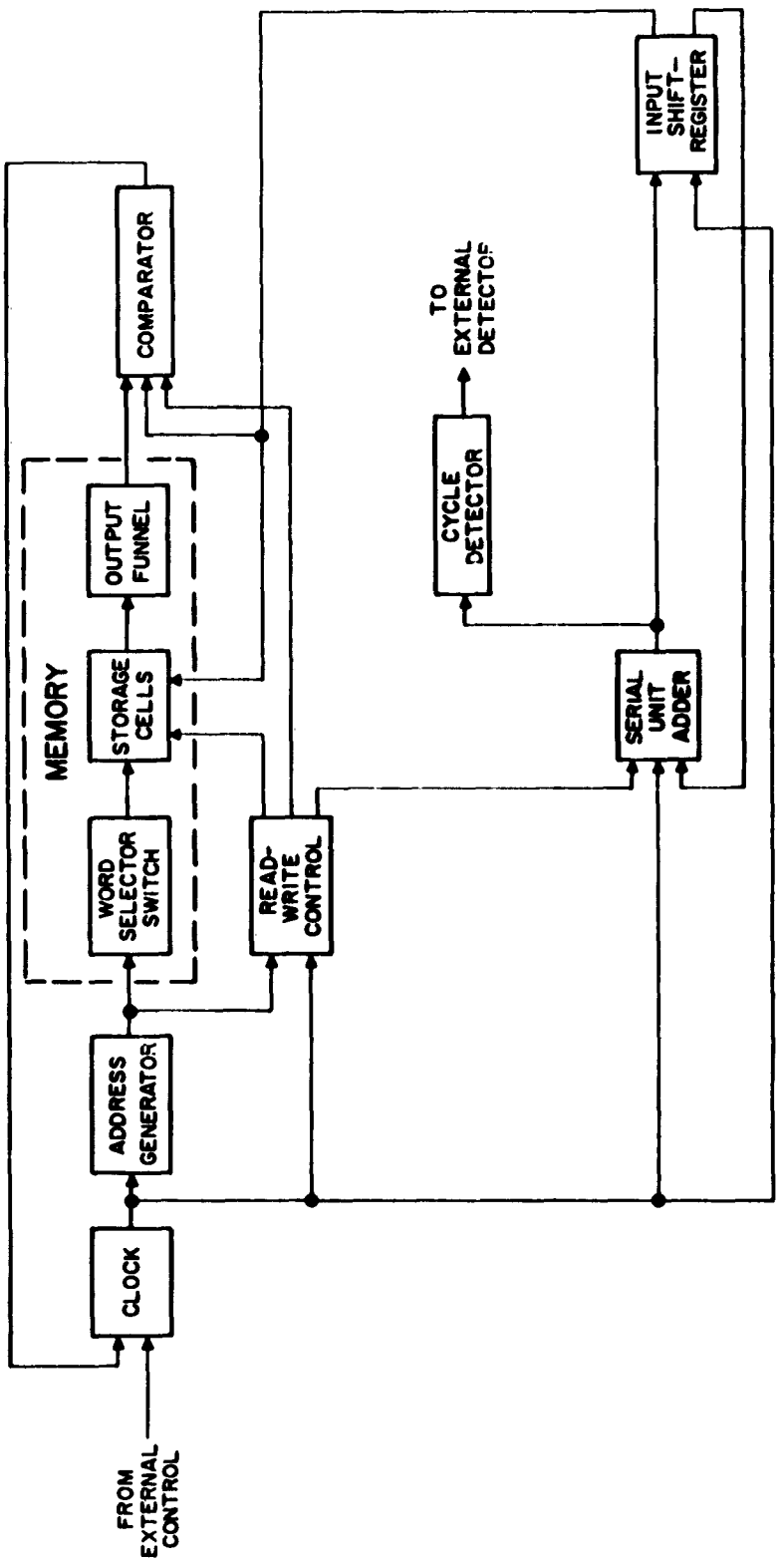


Figure 1 - Memory and control block diagram.

control, continuously at a rate determined by an external signal, or continuously under control of its own internal free-running cryotron-constructed circuits.

The detailed logic and cryotron circuitry for the memory control system are described in drawings that begin with condensed logical diagrams, proceed to detailed circuit diagrams for subunits of the system, and end with an integrated circuit drawing for the entire control system. The description of the circuits is followed by a discussion of the physical layout for the memory control system. This is followed by a discussion of the operation times for the system and its parts. Finally, some of the factors important in the design of high-speed cryotron circuits are analyzed.

SYSTEM SYNCHRONIZATION

This control system for the memory proceeds from one state to the next in the following way. Assume that the system has been halted so that all circuits have settled to a steady-state condition. In this halted condition the state of the memory control system is defined by the presence or absence of currents in the branches of its circuits. The next state of the system is determined by the present state of the system and the logic of the gates of the system.

The state of the system is only allowed to change when certain signals called clock signals are present. The clock pulse generator produces two non-overlapping control signals. One signal is designated phase 1 with the label C1, and the other signal is designated phase 2 with the label C2. Each of the two clock phases is fanned out from its source through repeating amplifiers so that all the gates that are clocked with the same phase receive their control signals as nearly simultaneously as possible. These phase signals are used in such a way that each signal may have a duration that can be varied from an indefinitely long time down to a minimum that is determined by the logic of the system and the operating times of the various circuits in the system. Each phase signal is usually introduced into the logic of the memory control system as one of two input control signals to a pair of AND-gates. One of the pair of AND-gates controls the current in one branch of a two-branch cryotron-circuit loop and the other AND-gate controls the current in the other branch of the same circuit. Thus, these clock-controlled (clocked) circuit loops cannot change state unless the proper phase of the clock signal is present. The primary objective of the non-overlapping characteristic of the clock phases is to insure that the system will pass through a proper sequence of states, independent of the duration of either clock phase signal and independent of the duration of the absence of both phases.

Thus either phase of the clock signal may remain on indefinitely or both phases of the clock may remain off indefinitely. The lower bound on the duration of the clock signals is set by the logic of the system and the operating time of the cryotron circuits. The minimum time between the beginning of phase 1 and the beginning of the next phase 2 is determined by the maximum circuit-path transit time from gates clocked with C2 to gates clocked with C1. The minimum width of each clock phase signal is less than the interval between the beginnings of successive phases. Each phase signal must last long enough to completely switch all the circuit loops over which it has direct control. This means that the clock signal does not need to remain on while unclocked circuit loops that are controlled by clocked loops are being switched. This type of clocking system was chosen in order to provide wide latitude in the testing of the entire system or any of its parts since it is expected that the circuit design constants, circuit layout, and circuit operating conditions will vary from one test situation to another.

SYSTEM OPERATION

General

The complete memory test operation consists of an alternating sequence of write cycles and read cycles. Each write cycle consists of 16 writing operations and each read cycle consists of 16 read operations. Each write and read operation is divided into two parts. One part is controlled by phase 1 (C1) and the other part is controlled by phase 2 (C2).

During phase 1 of a write operation, one bit is transferred from the 16-bit shift-register to one of 16 memory cells, and a 4-bit address is transferred from the address generator to the memory word selector to choose the proper memory cell. During phase 2 of the writing operation the address generator resets the circuits of the memory, a unit is added to the address in the address generator, and the shift-register is shifted one bit position. The 16 bits in the shift-register are transferred to the 16 memory cells during a write cycle. At the end of the write cycle the 16-bit pattern in the memory should be identical with the 16-bit pattern in the shift-register.

During phase 1 of a read operation a 4-bit address from the address generator is transferred to the memory and the selected bit is read out through the memory output funnel to the comparator.

Simultaneously, the corresponding bit is transferred from the shift-register to the comparator. The comparator generates an error signal if the two bits are unequal. This can be used to stop the system. During phase 2 of the read operation the address generator resets the circuits of the memory, a unit is added to the address in the address generator, and the shift-register is shifted one bit position. Since reading the memory is nondestructive, the 16-bit pattern stored in the memory remains identical with the pattern in the shift-register throughout the read cycle.

The address generator continuously cycles through 16 different 4-bit addresses and the shift-register shifts cyclically in a loop. However, the shift-register loop passes through a unit adder. The adder starts the addition of a unit to the 16-bit number in the shift-register at the beginning of the writing cycle and at no other time. A carry can be stored indefinitely between writing operations.

Notation of Logical Diagrams

The logic of the memory-control system that is used in writing is shown in Fig. 2. The notation[1] used on logical diagrams, such as this, is expressed by the interconnections among boxes that represent two types of logical elements. One type of logical element,

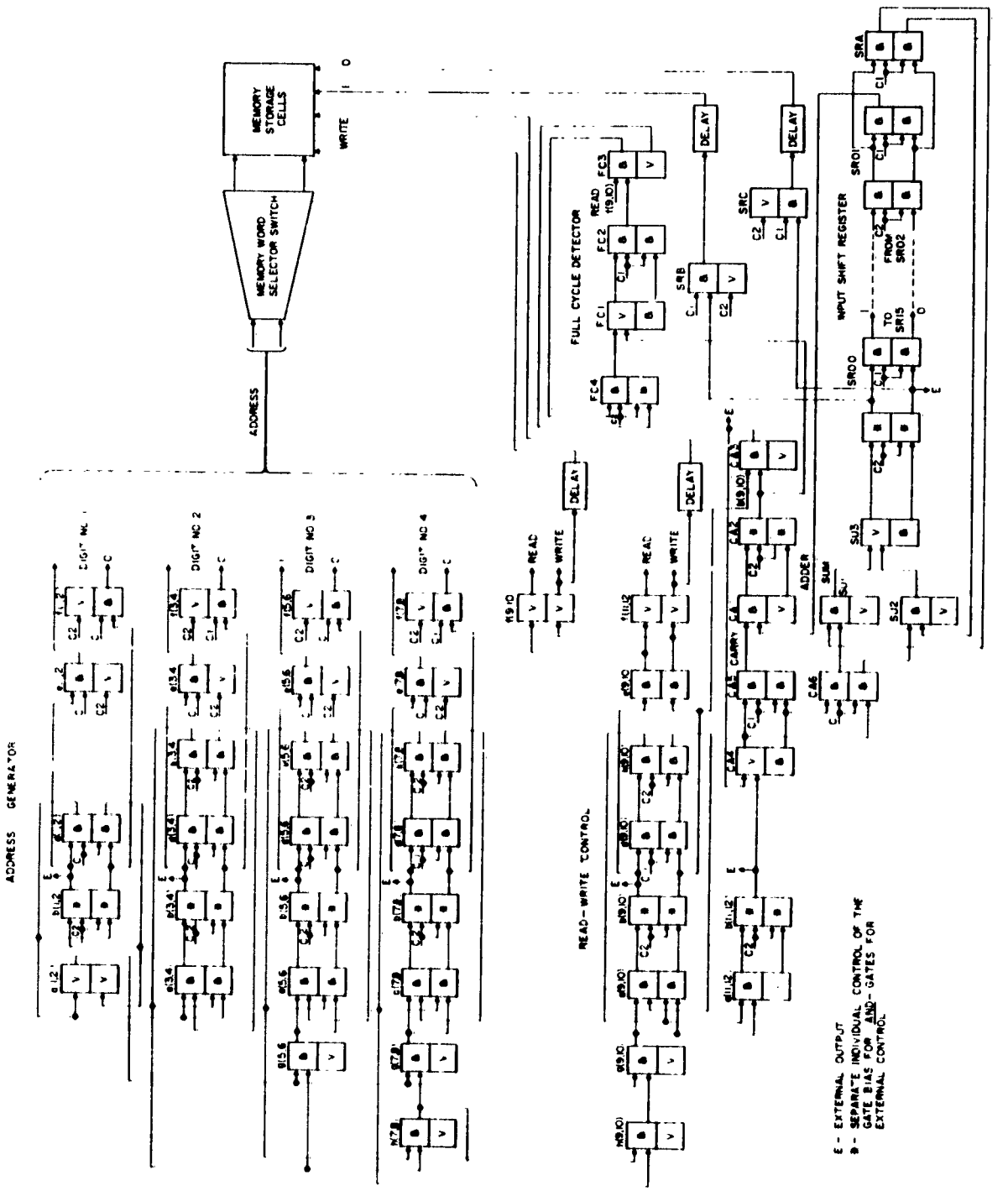


Figure 2 - Writing control logic.

the AND-gate, is represented by a box that contains the symbol &. The other type of logical element, the OR-gate, is represented by a box that contains the symbol V. When the circuits that are represented by the logic are in a steady-state condition, all logical gates are considered to be either on or off. AND-gates are on when all their inputs are on. OR-gates are on when any one or more of their inputs are on. Two input AND-gates and two input OR-gates can be constructed out of in-line cryotron gates with two levels of control. Two-branch cryotron circuits that have one two-level control gate in each branch can be represented on the logical diagrams by a pair of boxes, one above the other with two inputs per box. Since these two-branch cryotron circuits must not have resistive gates in both branches simultaneously (when they are in the steady-state condition) the corresponding pairs of logical elements should never be on simultaneously. For this reason, many pairs of logical elements are designed so that they operate as logical duals of one another. For example, if an AND-gate has inputs A and B, the associated logical element is an OR-gate with inputs $\overline{A}$ and $\overline{B}$. Certain AND-gates are paired together with a common clock signal (C1 or C2). The second input to a clocked gate is generally the logical dual of the second input to the associated clocked gate. In order that the logical notation should not be misleading the following cryotron-circuit operating conditions should be

noted: A cryotron AND-gate becomes resistive when current flows in both of its input control lines in a direction opposite to the direction of the gate current. The output current for this circuit is forced to flow in the other branch (not containing this AND-gate) of the two-branch circuit. The output current will remain in the other branch even after the input currents to the AND-gate disappear unless it is forced back into the branch containing the AND-gate by resistance in a second gate in the circuit. The logical diagrams have been simplified by leaving out the input and output connections of boxes that represent the second half of a circuit that is to be formed from a pair of logically dualized gates. This is done whenever the addition of these lines does not aid in the understanding of the logic of the system. Each pair of boxes on the logical diagram can be considered to be in the 0-state or the 1-state depending upon which of the two boxes is on. An arbitrary assignment of a correspondence between a particular state and a particular logical gate condition is made at certain key pairs of gates in a system and the correspondence between states and gate conditions for the remaining gates follows from these. The logical diagrams are arranged in most cases so that the pairs of boxes are considered to be in the 1-state when output from the upper box is on and the output of the lower box is off, and in the 0-state when the output from the lower box is on and the output from the upper box is off. The two-

branch circuit that corresponds to a pair of logical gating boxes is considered to be in the 1-state when current is flowing in the branch that corresponds to the 1-state output line of a pair of logical gating boxes.

The Logic of the Address Generator

The address generator is a binary counter. It is represented by the four rows of pairs of gates in the upper left portion of Fig. 2. The logic of digit 1, the least significant digit, is represented by the top row of pairs of gates. The pair of AND-gates b1 and b2 in the second column from the left in this top row, clocked with C2, and the pair of AND-gates d1 and d2 in the third column from the left in this top row, clocked with C1, represent the storage and inverting logic for digit 1. During phase 1 (when C1 is on and C2 is off) information is transferred from the gates b1 and b2 in column two to the gates d1 and d2 in column three, while the gates b1 and b2 in column two remain unchanged. During phase 2 (when C2 is on and C1 is off) the dual of the information stored in the gates d1 and d2 in column three is transferred into the gates b1 and b2 of column two while the gates d1 and d2 in column three remain unchanged. It is assumed, as indicated in Fig. 2, that the 1-state of any of the pairs of gates of digit 1 is

represented by the top gate's being on and the bottom gates's being off. The 0-state in one of the pairs of gates is represented by the inverse condition. The pair of gates e1 and e2 in column 4 of digit 1 functions as the 1-state driver stage for the memory word selector. The 1-state is represented by the top gate's being on. The pair of gates f1 and f2 in column 5 of digit 1 functions as the 0-state driver stage for the memory word selector. The 0-state is represented by the bottom gate's being on. To insure that the 0-state and the 1-state of digit 1 are never simultaneously sent to the memory, the upper gate in column four and the lower gate in column five are clocked with C1 while the other gate in both columns is clocked with C2. Thus C1 represents the memory address setting time period and C2 represents the memory address resetting time period. The leftmost pair of gates a1 and a2 for digit 1 functions as repeaters for the pair of gates d1 and d2 of column three and represents the transmission of the carry initiated by digit 1 to the other digits of the address generator. The pairs of gates beneath those of digit 1 in columns two, three, four, and five provide the equivalent functions for digits 2, 3, and 4. The pairs of gates on the left end of each row for these other digits provide the gating for their carry inputs from less significant digits. The pairs of gates labeled k3 through k8 for digits 2, 3, and 4 prevent a carry from changing these digits

a second time in the event that the disappearance of the carry signal is delayed into the pairs of gates labeled a3 through a8.

It is desirable to reduce the propagation of the carry from digit 1 to digit 4 so that the rate at which the address generator can produce successive sets of 4-digit addresses can be increased. This is accomplished by combining the value of digit 1 with the values of the other digits that form a particular carry after the values of the more significant digits have been combined. If the carry from digit 1 to digit 4 were generated serially, the carry circuits would require six gate pairs (six circuit loops), but the carry-propagation time during C1 would be five loop-switching times. Instead, this design utilizes seven gate pairs (seven circuit loops) and the carry-propagation time during C1 is four loop-switching times. However, one of the loops has a fan-out of six, while all other loops have a fan-out of four.

Several alternative designs were considered for the address generator. On the basis of speed alone, the fastest address generator would be a 16-bit shift-register ring, such as the one used for information storage. This would circulate one of several possible 16-bit patterns. It could generate addresses at a rate of one every two loop-switching times. Each loop in the shift-register design has a fan-out of two. The carry propagation time during C1 (equivalent in function

to the propagation of a carry from digit 1 to digit 4 in the binary counter design) is reduced to a single loop-switching time. This type of address generator was not used since it required about twice as many gates as the binary counter. The input shift-register could serve the double function of address generation and memory-input-information generation, but this would severely limit the set of patterns to be written into the memory. It should be noted that the sequence of addresses generated by a shift-register would not follow the binary number sequence although this does not constitute an objection to its use for this control system. One 16-bit pattern that would sequence through 16 addresses is

0000110100101111.

Other alternative address-generator designs that compromise between the number of gates required and the speed of address generation are based on combinations of shifting rings containing fixed patterns. One example of a compromise is the use of two 4-bit rings each containing the pattern 0011. One of the two rings would shift at a maximum rate, the other would shift only every fourth shift of the first. The first ring would produce the two less significant address digits, the second ring would produce the two more significant address digits. The signal that controls the shifting of the second ring is based on an examination of a pair of digits in the first ring, such as the

occurrence of a pair of ones in two adjacent stages. This signal must then fan out to shift all four stages of the second ring. The time required to fan out the shifting signal is equivalent to a carry propagation time.

The Logic of the Read-Write Control

The read-write control flip-flop is located on Fig. 2 in the row of gates just below the address generator. This flip-flop functions exactly as a fifth address digit of the address generator would function. It changes state from read to write or from write to read whenever a carry is propagated to it from digit 4 of the address generator.

The Logic of the Shift-Register and Adder

The input shift-register is located in Fig. 2 in the lower right. This shift-register has 16 stages but only two are shown; SR00 and SR01. Each stage consists of two pairs of gates. The input pair of gates of a stage is clocked with C2. The output pair of gates of a stage is clocked with C1. It is assumed, on Fig. 2, that a stage is in the 1-state when the top gate of a pair is on, and that a stage is in the 0-state when the bottom gate is on. During phase 1,

information contained in the pair of input gates of each stage is transmitted to its pair of output gates. During this time the pair of input gates clocked with C2, which is absent, is not allowed to change. During phase 2, information contained in the pair of output gates of each stage is transmitted to the adjacent stage. Information in the output half of stage SR00 is transmitted to the input half of stage SR15, information in the output half of stage SR15 is transmitted to the input half of stage SR14, etc., and information in the output half of stage SR01 is transmitted to the input half of stage SR00. The path from stage SR01 to stage SR00 passes through the summing gates, SU1 or SU2 and SU3 of the unit adder. If a carry had been generated from the previous digit that passed through the adder, the summing gates invert the value of the digit that is being transmitted from SR01 to SR00. If no carry had been generated from the preceding digit the current digit is transmitted from SR01 to SR00 unchanged. The three pairs of gates that generate the sum for the adder are not clocked. Thus the time interval between the initiation of C1 and the initiation of C2 must be sufficient for the information in SR01 to be transmitted through the sum gates of the adder to the input gates for stage SR00. The adder-carry stage consists of the pairs of gates labeled CA1 through CA5. These are in the lower center of Fig. 2 just above the summing gates. Again a pair of gates

is assumed to be in the 1-state when the upper gate of the pair is on, and in the 0-state when the lower gate is on. The carry stage is initially turned on by the propagation of a carry signal from digit 4 of the address generator by means of gates a12 and b12 when the read-write flip-flop is in the read state. The carry flip-flop will be kept on as long as the 1-state of stage SR01 of the shift-register is maintained or until the write signal from the read-write flip-flop disappears. The read-write flip-flop is changed from read to write by the same signal, g10, that initially turns on the carry stage.

The effect of the unit adder is to add one unit to the number in the shift-register during each cycle of 16 writing operations. Since the number in the shift-register is modified only by this unit addition, the number increases from 0 to $2^{16}-1$ and then returns to 0. Each of these patterns of 16 bits is written into the memory and then read out and checked.

System Monitoring and Manual Control

The full-cycle detector is a flip-flop that consists of the pairs of gates labeled FC1 through FC4. It is turned on by the 1-state of the carry stage at the end of a write operation and turned off by the end of the next read operation. The memory

control system will sequence through 16 write and 16 read operations for each pattern in the shift-register, and the shift-register will sequence through 2^{16} patterns. Each read and write operation has two phases. Thus the control system will sequence through 2^{22} states before it repeats a state. When the system is running continuously the full-cycle detector will be turned on once in each complete control sequence and it will stay on for the next 16 read operations. Its function is to provide a signal for an outside observer to serve as a gross indication that the system is running and the rate at which it is running.

Externally controlled means of clearing the control system to a particular state (i.e., forcing any or all storage elements to either the 0-state or 1-state) are to be accomplished by individual external control over the biases of both gates in certain selected circuit loops.

External observation of the state of a selected set of storage elements of the memory-control system is to be accomplished by driving buffer-cryotron gates from the output of the storage elements. These buffer cryotrons would not need to switch at a fast rate since they do not feed their outputs back into the system, and their outputs are only intended to be observed externally when the control system has been halted and the circuits have settled to a steady-state condition.

The Logic of the Comparator

There are two models of the memory that operate as follows during reading. Memory Model A has an output funnel that terminates in a single circuit loop. Current in one branch of this output circuit loop indicates that the memory storage cell being read is in the 1-state. However, current in the other branch of this output circuit may indicate that the storage cell is in the 0-state or that the memory has been reset. Memory Model B has an output funnel that terminates in a pair of circuit loops. One of these loops has one branch that contains current when a storage cell in the 1-state is being read. The other loop has one branch that contains current when a storage cell in the 0-state is being read. The second branch of each of these loops contains current when the memory has been reset.

The logic of the control system that is intended to operate with Model A of the memory is shown in Fig. 3a. This figure does not repeat the detailed logic shown on Fig. 2 of subunits of the memory control that function during reading in the same way that they function during writing. The comparator consists of pairs of gates labeled CO1 through CO4. It forms the unequals function of a bit from the memory and a bit from the shift-register. The comparator used with Model A of the memory forms the unequals of the

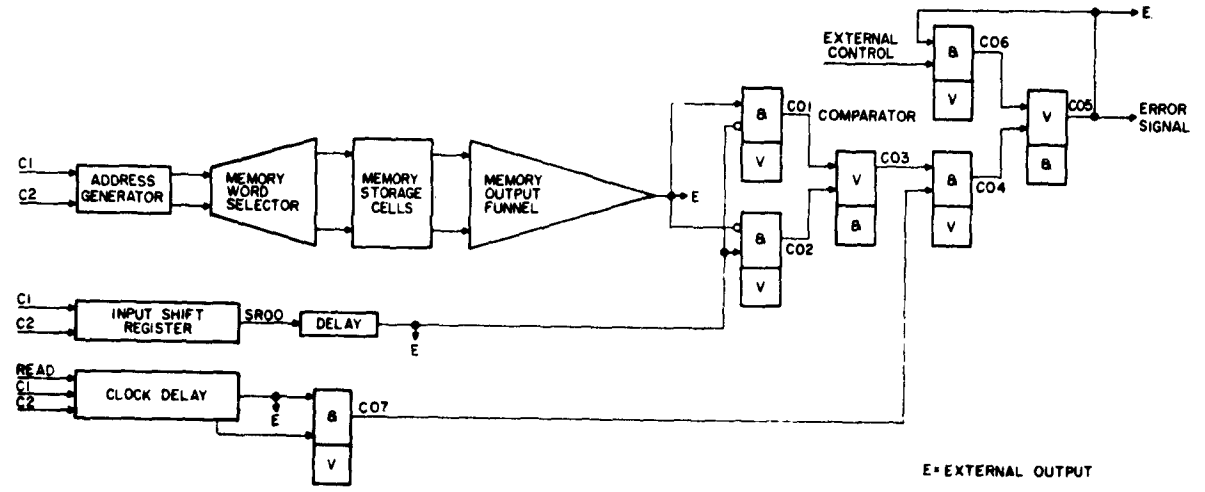


Figure 3a - Reading control logic (Model A memory).

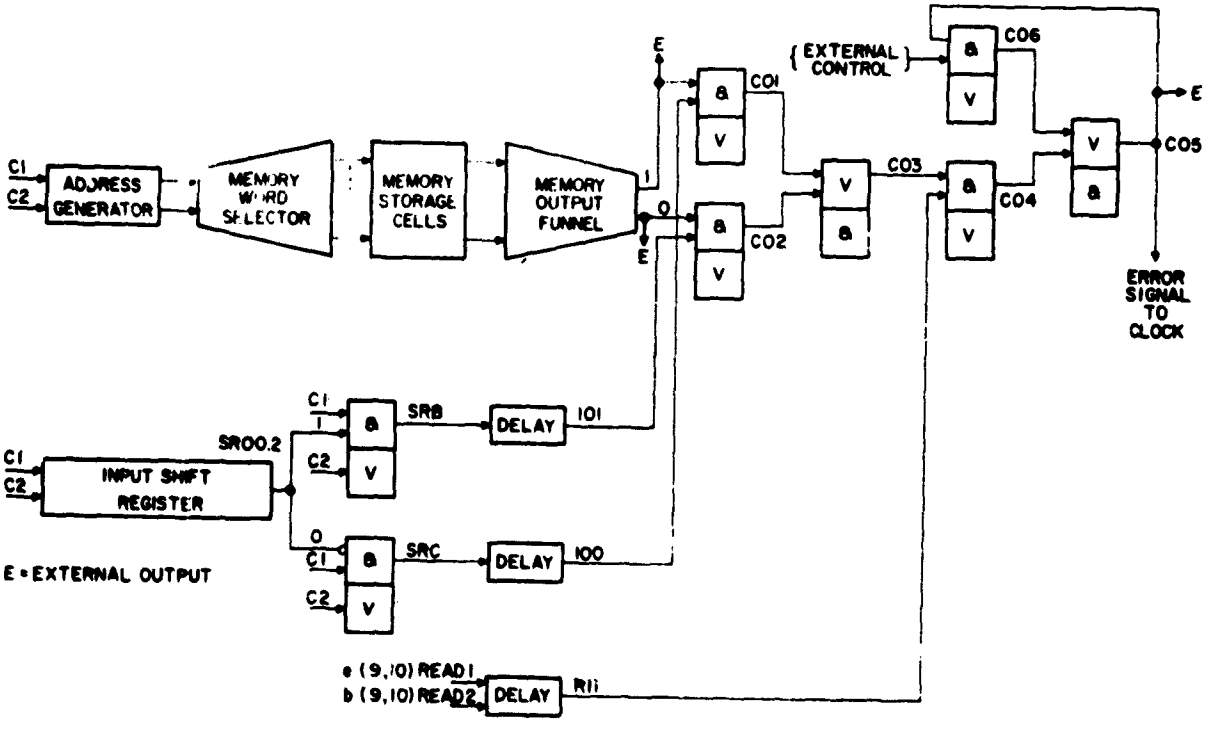


Figure 3b - Reading control logic (Model B memory).

single-circuit output of the memory and the single circuit output of a delay line from the shift-register. In order to prevent the indication of false errors a delayed and narrowed read signal clocked with C1 gates the signals from the output of the unequals circuits into error storage circuits. Model B of the memory produces a 1-circuit output and a separate 0-circuit output, and both circuits are in a reset condition during phase 2. Accompanying these Model B signals is a separate 0-circuit signal and a 1-circuit signal, present only in phase 1 and transmitted from the shift-register through delay lines to the comparator. Figure 3b shows the logic of the test system for the Model B Memory. The error signal can be used, under the option of external control, to stop the generation of clock signals and halt operation. The error signal is also transmitted outside the refrigerator for external error indication.

The maximum clock-signal repetition rate is determined by two factors for each clock phase: first, the circuit loops in the system that have the longest switching time, and second, the greatest time difference between any pair of signal paths that start in circuits clocked with one of the phases, pass through unclocked circuits, and finally terminate in circuits clocked with the other phase. The signal transit time through the circuit path from address generator to memory storage cells is longer than the signal transit time through

the circuit paths from the read-write flip-flop to memory storage cells and from the input shift-register to memory storage cells. Thus special delay lines are added in the latter paths so that these differences do not slow down the clock repetition during writing operations. The signal transit times through the circuits paths from address generator through the memory to the comparator, from the read-write flip-flop to the comparator, and from the shift-register to the comparator are equalized by special delay lines in order to increase the clock repetition rate during reading operations. Since there is a longer natural delay in the memory circuit path to the comparator, special delay lines are used in the other paths to the comparator. These delay circuits require a switching rate capability that is greater than the clock repetition rate if they are to be effective. This can be achieved by making the delay line from a series of two-branch cryotron circuits, each circuit having a fan-out of two. These delay lines are not needed to make the system work but only to allow the system to work at a faster rate.

THE IN-LINE CRYOTRON CIRCUITS

General

This section describes separate circuit diagrams for typical portions of each of the subunits of the memory control system. An integrated circuit layout shows how the various subunits are arranged physically relative to one another and relative to the memory. The individual circuit diagrams of the subunits differ from the same circuit in the integrated system layout in the following respects:

1. The source of the clock signals and their proper fan-out is not indicated on the individual diagrams.
2. The gates in the individual diagrams are positioned so that the logic can be interpreted more easily.
3. The spacing between gates in the individual diagrams is uniform and does not allow proper space for wires that pass between the gates.
4. The interconnections among gates are not routed optimally.
5. The current sources and sinks for the various circuit loops are not interconnected.

A considerable additional effort in the integrated circuit layout has been expended, first to position the gates and route the intergate connections so that the system as a whole will operate at greater speed,

and then to connect the current sources and sinks of the various circuit loops in series in several separate chains so that testing for short circuits will be possible.

Circuit Notation

The basic circuit notation[1] is the same on all circuit diagrams. Figure 4 may be examined as typical. Each gate is represented by a rectangle. Each circuit is a loop with a dc source and a dc sink, and is divided into two branches. Each branch of a circuit loop contains one gate. All the lines on the diagrams except the borders of the rectangles represent the paths of wires that join pairs of gates together or supply the source or sink current to the loops. Each of the lines is in reality a wire that has a width equal to the width of a gate. The physical circuits represented by these diagrams will provide insulation between the wires in different circuit loops. The width of the separation between wires that lie in the same plane is presently considered to be equal to the width of the wiring. The insulation between control wires and gates and between intergate wires that cross one another is much thinner than the wire width. The complete wiring loop of each circuit consists of two gates represented by two rectangles and the wire interconnecting

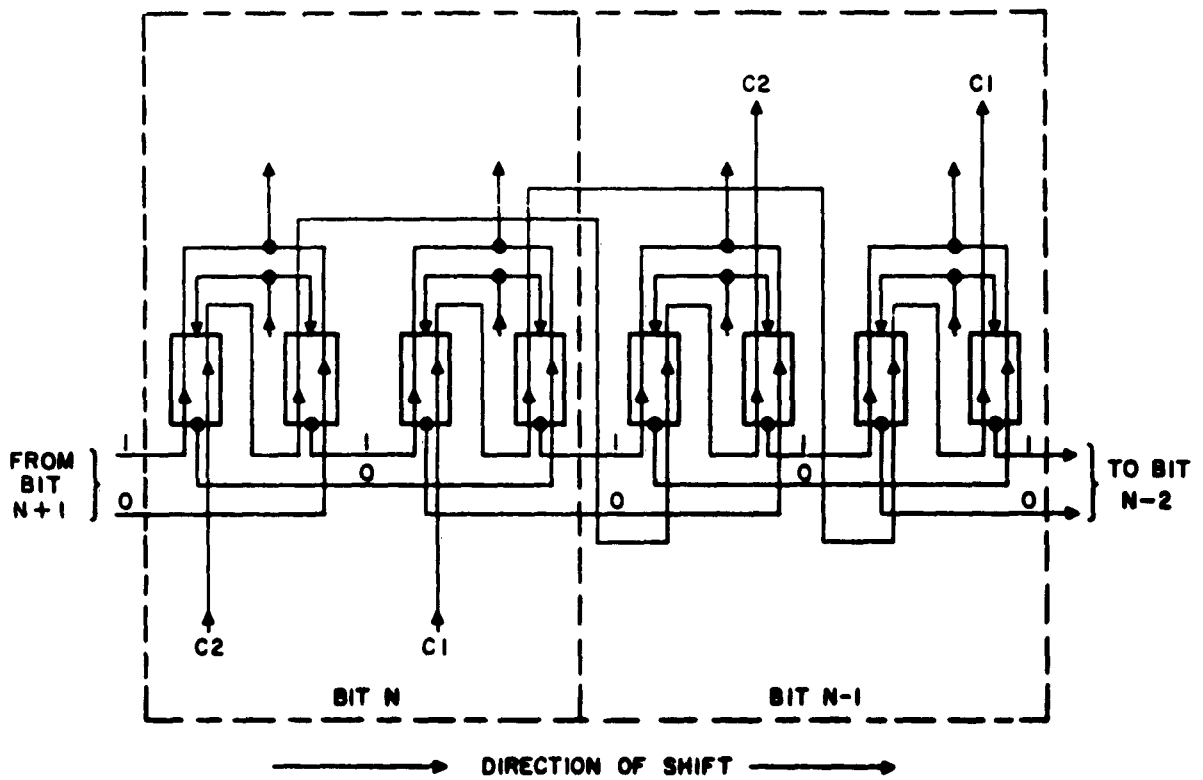


Figure 4 - Circuit for two bits of a shift-register.

them represented by the lines connected to the center of each end of the rectangles. An arrowhead into one end of each rectangle indicates the direction of flow of the current through the gate. The control wire inputs to the gates are represented on the circuit diagrams by the lines passing through the rectangles. These wires, equal in width to the gate, pass over the gate in one of two levels. The inductance of the control wire closer to the gate (the lower level) is less than that in the control wire farther from the gate. Thus, the control wires of circuits that must switch most rapidly because of their system function are routed to the lower level. The arrowheads inside the rectangle indicate the direction of current flow in the control lines and also indicate the level intended for each control wire. The control line with an arrowhead closer to the arrowhead that indicates the direction of current flow through the gate is intended to indicate the lower level control line. When the direction of current flow in a control line is opposite to the direction of current in the gate, the control line represents a logically positive input. When the direction of current flow in a control is in the same direction as the current in the gate, the control line represents a logically negative input (an inhibit signal line). In the case of an inhibit line a circle is placed on the control line inside the rectangle at the end that has a gate current arrowhead. All rectangles imply AND-gate circuits

except those rectangles that contain special symbols. A rectangle that contains a plus (+) sign implies an OR-gate circuit. A rectangle that contains an asterisk (*) indicates an AND-gate in which the bias for that gate is to be under separate external control. Circled plus signs ⊕ indicate separate external bias control for an OR-gate. AND-gate circuits are biased so that two positive control currents are required to switch the gate resistive. OR-gate circuits are biased so that one or two positive control currents will switch the gate resistive. An additional unit of bias current is required for each negative control current input to a gate over that needed if all inputs are positive.

The Shift-Register Circuits

Figure 4 shows two bit-storage stages of a shift-register identical with those used in the input shift-register. The information shifts from left to right. Each of the stages contains two circuit loops with two gates apiece. The input circuit of each stage is clocked with C2 and the output circuit is clocked with C1. Assume that both C1 and C2 are off (neither clock signal is carrying current) and that all the circuits have reached a steady-state condition with all gates nonresistive. A dc signal is flowing in one and only one of the two branches of every circuit loop. Assume that N, the circuit

designation symbol on Fig. 4, is even and that current is flowing in the 1-digit side of the circuits of all even-numbered stages and that current is flowing in the 0-digit side of the circuits of odd-numbered stages. Then, if the clock signal C1 comes on, the left-hand output gate of stage N and the right-hand output gate of stage N-1 go resistive. No current is switched in the circuits, however, since there is no current flowing through the gates that become resistive. Next, let C1 go off and after it has gone off let C2 come on. Under these conditions, the right-hand input gate of stage N and the left-hand input gate of stage N-1 go resistive. In this case, the current that has been flowing in these branches of the input circuits will be driven over into the other branch of these circuits. Thus, the input circuit loop of stage N is switched from the 1-state to the 0-state and the input circuit loop of stage N-1 is switched from the 0-state to the 1-state. During the next C1 signal, the output circuits of all stages will change state.

The Address Generator Circuits

Figure 5 shows two digit stages of a binary address generator similar to those used in the address generator of the memory control system. These stages are intermediate stages of a binary counter in

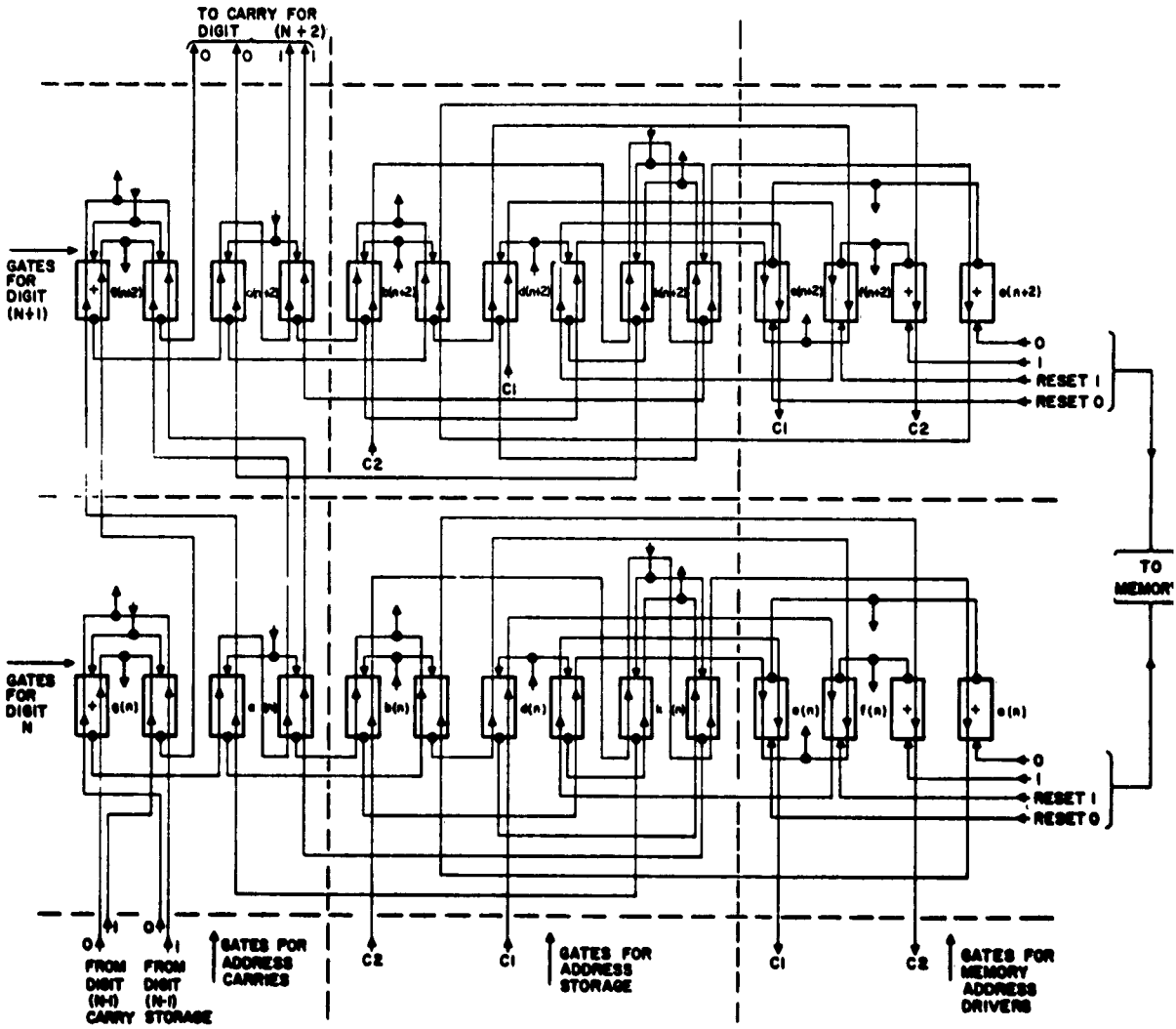


Figure 5 - Circuit for two digits of an address generator.

which each more significant digit stage of a pair, such as those indicated, changes state at half the frequency of the less significant stage. The gates associated with the less significant digit N are in the bottom row of Fig. 5. The gates associated with the more significant digit N+1 are in the top row of Fig. 5. The logic of the connections for each gate of digit N is identical with the logic of the connections for the corresponding gate of digit N+1. All digit stages, except the first, of a binary counter would be identical with the two that are on Fig. 5. The initial stage of the counter would be simplified since no carry comes into it. The carries indicated in Fig. 5 are propagated serially through the carry circuits of successive stages of digits. It has been pointed out that some speed can be gained by distributing the carry from digit 1 in a more direct route to the other digits. Gates that generate the carry signals are shown in the four left-hand columns of gates on Fig. 5. The address digit storage is maintained in the six middle columns of gates. The gates that alternately drive the memory-word selector, first with address-setting signals and then with memory-resetting signals, are shown in the four right-hand columns of gates.

The address-storage gates are divided into three parts. The left-hand pair of the six address-storage gates in each row forms an input storage circuit clocked with C2. The middle pair of gates forms

an output storage circuit clocked with C1. The right-hand pair of gates, clocked with C2, prevents hazardous switching of the right-hand pair of carry gates. During C1, the output storage circuit for each digit is allowed to switch to the opposite state and is made to agree with the state of the associated input storage circuit. The input storage circuits for each digit are held in their current state by the absence of C2. When the output storage circuit of digit 1, the least significant digit, changes from the 0-state to the 1-state, the change will propagate during C1 through the address carry gates. The state of successive carry stages will change up to and including the carry input circuit for the first digit state that is in the 0-state. Also, during C1 one of the pair of memory address driver circuits, the right-hand side of Fig. 5, is switched from the reset state to the 1-state or the 0-state that is indicated by the associated input storage circuit. The resulting address is transmitted during C1 to the memory word selector. During C2 the input storage circuit for each digit is allowed to switch to the opposite state and is made to be equal to the dual of the state of the associated output storage circuit if a carry has been propagated to it from the next less significant digit stage. The output storage circuits for each digit are held in their current state by the absence of C1. Also during C2 one of the pair of memory address driver circuits for each digit is switched from the previous "1" or "0"

state to the reset state. Thus, all memory address driver circuits end up in the reset state at the end of phase 2.

The Adder and Full-Cycle Detector Circuits

The adder, its connections from the read-write flip-flop and shift-register stage SR01, and its connections to the full-cycle detector and shift-register stage SR00 are shown on Fig. 6. The pairs of gates that form the adder carry storage loop have been labeled CA1 through CA6. The pairs of gates that form the adder sum have been labeled SU1, SU2, and SU3. The pairs of gates that form the full-cycle detector have been labeled FC1 through FC4. The two pairs of gates that constitute stage SR00 of the shift-register have been labeled SR00. The circuits formed from pairs of gates that contain asterisks, indicating special individual bias control, are capable of being put into either the 0-state or the 1-state by manual control of their bias. An initial state for the system would be set up when the clock phase signals are absent. Then, in preparation for continuous clocked operation, the bias of these gates would be adjusted so that they would operate as two-input AND-gates.

The adder carry storage loop is put into the 1-state by a signal from the address generator (the same signal that switches the

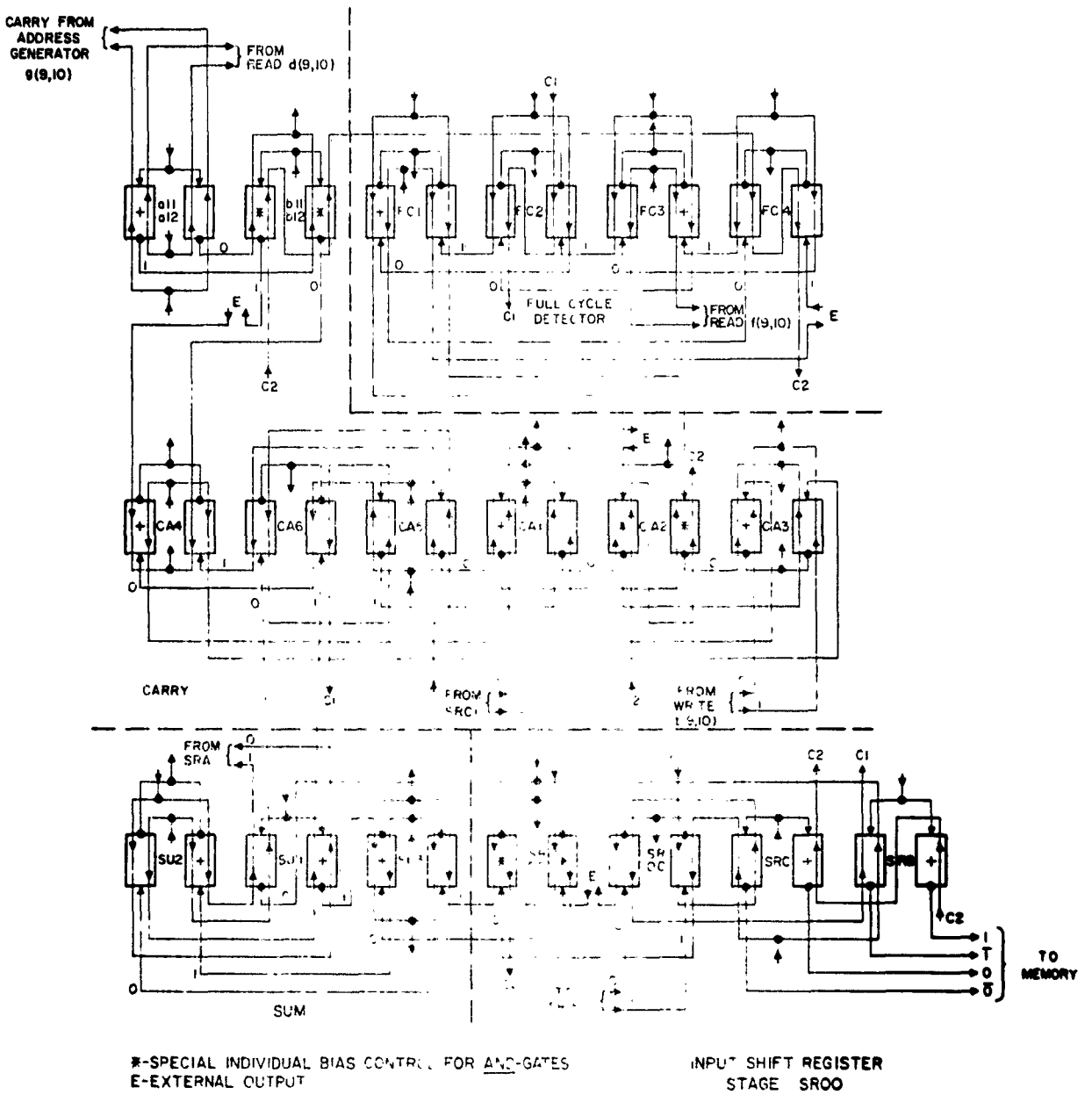


Figure 6 - Circuit for an adder and full-cycle detector.

read-write flip-flop) and a read signal AND-gated together into the a12 circuit during phase 1. Circuit a12 switches circuit b12 into the 1-state during the next C2. Then circuit b12 switches circuit CA4 into the 1-state. During the next C1, which coincides with the beginning of a write cycle, the CA4 switches the CA5 and CA6 circuits into the 1-state. The CA6 circuit is a repeater for CA5. If the shift-register stage SR01 is also in the 1-state, circuit CA5 AND-gated with circuit SR01 will switch circuit CA1 into the 1-state. This constitutes the generation of a carry bit to be propagated from the least significant digit of the number in the shift-register to the next more significant digit in the shift-register. During the next C2 signal, the CA2 circuit will be put in the state that corresponds to the state of the CA1 circuit. The CA3 circuit will be put in the same state if the write signal is on; otherwise the carry storage circuit will be cleared to the 0-state. The CA4 circuit which picked up the initial carry will transmit a carry that has been propagated to it from the CA3 circuit. The carry loop will continue in the 1-state until either a 0-state is indicated by stage SR01 or until the write signal disappears.

When the CA6 circuit is in the 0-state, the SU1 circuit of the adder transmits the 1-state of the SR01 stage of the shift-register through the SU3 circuit to the SR00 stage of the shift-register. When

the CA6 circuit is in the 1-state, the SU2 circuit of the adder transmits the dual of the 0-state of the SR01 stage of the shift-register through the SU3 circuit to the SR00 stage of the shift-register.

The full-cycle detector is considered to be on when stages FC3 and FC4 are in the 1-state. This is accomplished by a signal from CA2 switching circuit FC1 and then FC2 into the 1-state when the read signal is on.

The Delay Line and Comparator Circuits

Figure 7 shows the circuit for a variable length cryotron delay line. The length of the line is varied by selecting a particular input circuit for the line. Three such input circuits are represented by the three columns of pairs of gates on the left half of Fig. 7. The gates on the right half of Fig. 7 represent three simple repeater circuit loops for the delay line. An input circuit is selected by appropriate settings of the bias control currents for the gates on the bottom row of Fig. 7. The gates intended to be under external control are indicated by the symbol ⊕ inside the gate symbol. The input signal to the delay line is represented by the line labeled S and its dual labeled $\overline{S}$, which control the state of the bottom row of gates.

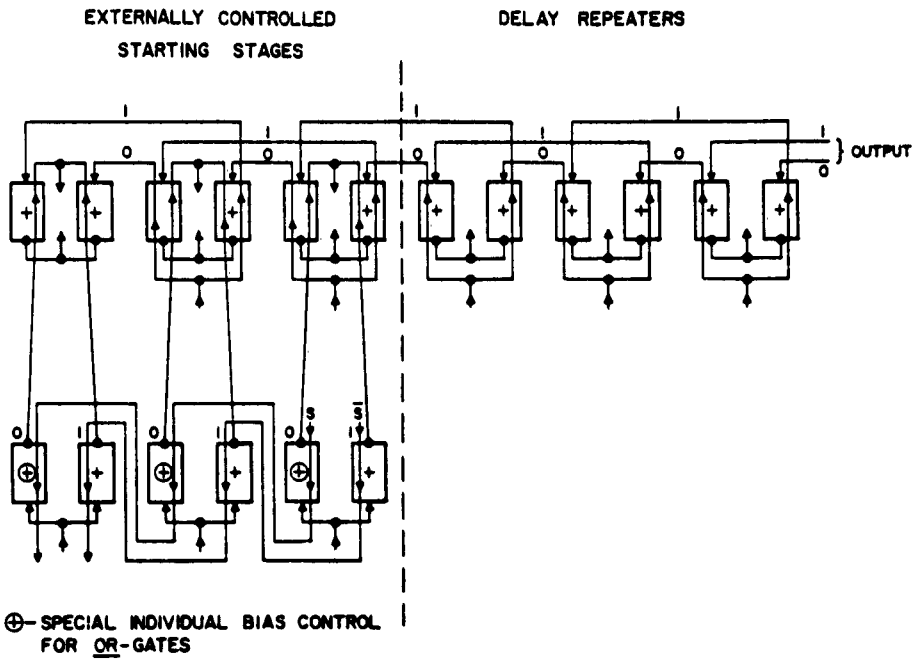


Figure 7 - Circuit for a variable length delay line.

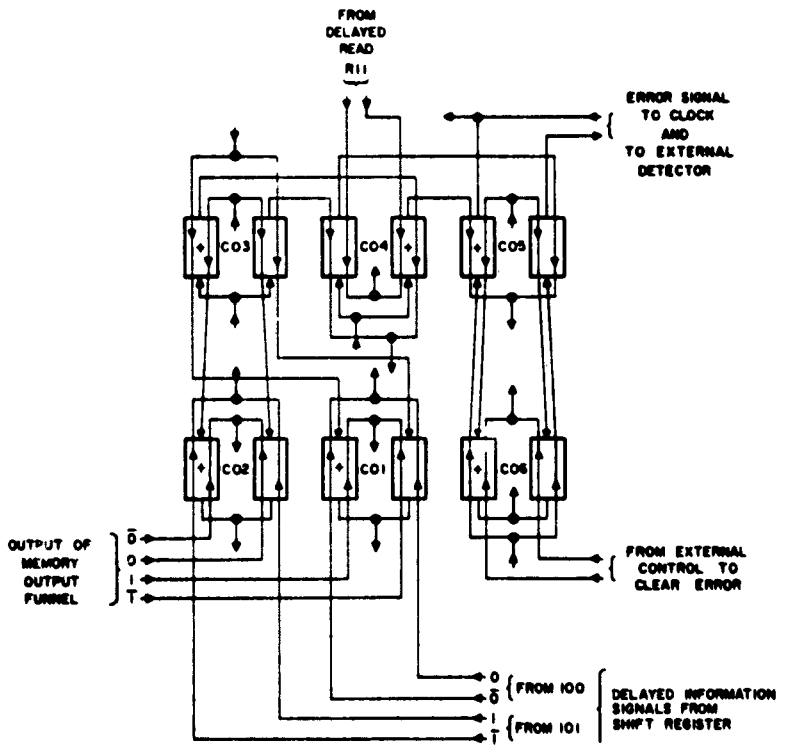


Figure 8 - Circuit for a comparator.

In order to operate this delay line with a minimum delay, the bias of the rightmost gate marked ⊕ would be adjusted so that the gate operates as an OR-gate. Thus, current on the S line would make the gate resistive. The other two gates marked ⊕ would be biased so that they could never go resistive. The minimum delay would be five loop switching times. For maximum delay of 7 loop switching times, the left-hand gate marked ⊕ would be biased as an OR-gate and the other two would be biased so that they could not go resistive. This circuit is used to equalize the delay from the shift-register to comparator and the delay from memory to comparator.

The circuits of the comparator are shown in Fig. 8. The circuits CO1, CO2, and CO3 generate a signal that is on (CO3 in the 1-state) when the value of the bit read from the memory is unequal to the value of the bit transmitted to the comparator from stage SR00 of the shift-register. Circuit CO4 transmits the unequal signal from circuit CO3 only when an appropriately delayed C1 and read signal are present. Circuits CO5 and CO6 will store error signals that are indicated by any 1-state of circuit CO4.

The Clock Pulse Generator Circuits

A logical diagram of the clock pulse generator is shown in Fig. 9 and its circuit is shown in Fig. 10. This circuit consists of a series of loops connected to form a ring. The number of loops in the ring may be varied from a minimum of four to a maximum of six. The 1-side of circuit KO1 is AND-gated with the 0-side of circuit KO2 in circuit KO9 to produce the C1 signal. The 0-side of KO1 is AND-gated with the 1-side of KO2 in circuit K10 to produce the C2 signal. An error signal which controls the KO3 circuit can stop the clock and thus halt the system. The external control over the bias currents for the gates of circuits KO4, KO5, and KO7 provides the means by which the clock pulse repetition rate can be varied between 8 and 12 loop switching times when the clock is free-running. The external control input for circuit KO7 provides a means of external determination of the clock pulse repetition rate. The pulse width of each phase of the clock is only adjustable to the extent that the switching time of the KO2 circuit can be varied by adjustment of its bias. The width of each clock pulse equals one loop switching time. By means of external control over the bias of circuit KO2, the ring may be stopped with either C1 or C2 on. The ring may be stopped with neither C1 nor C2 on by control of the bias of the circuit of KO7.

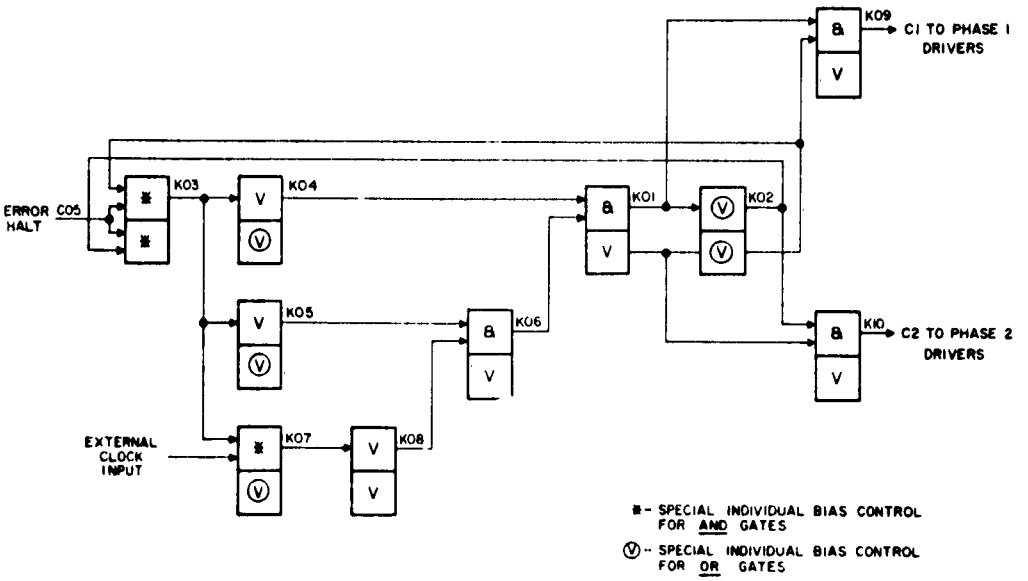


Figure 9 - Clock pulse generator logic.

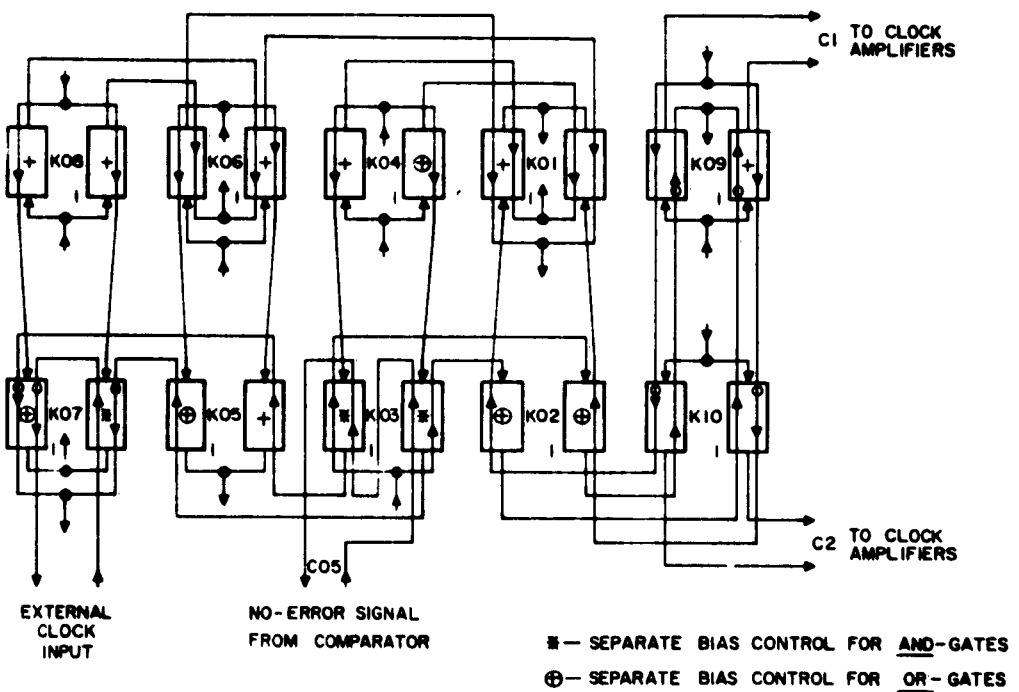


Figure 10 - Circuit for a clock pulse generator.

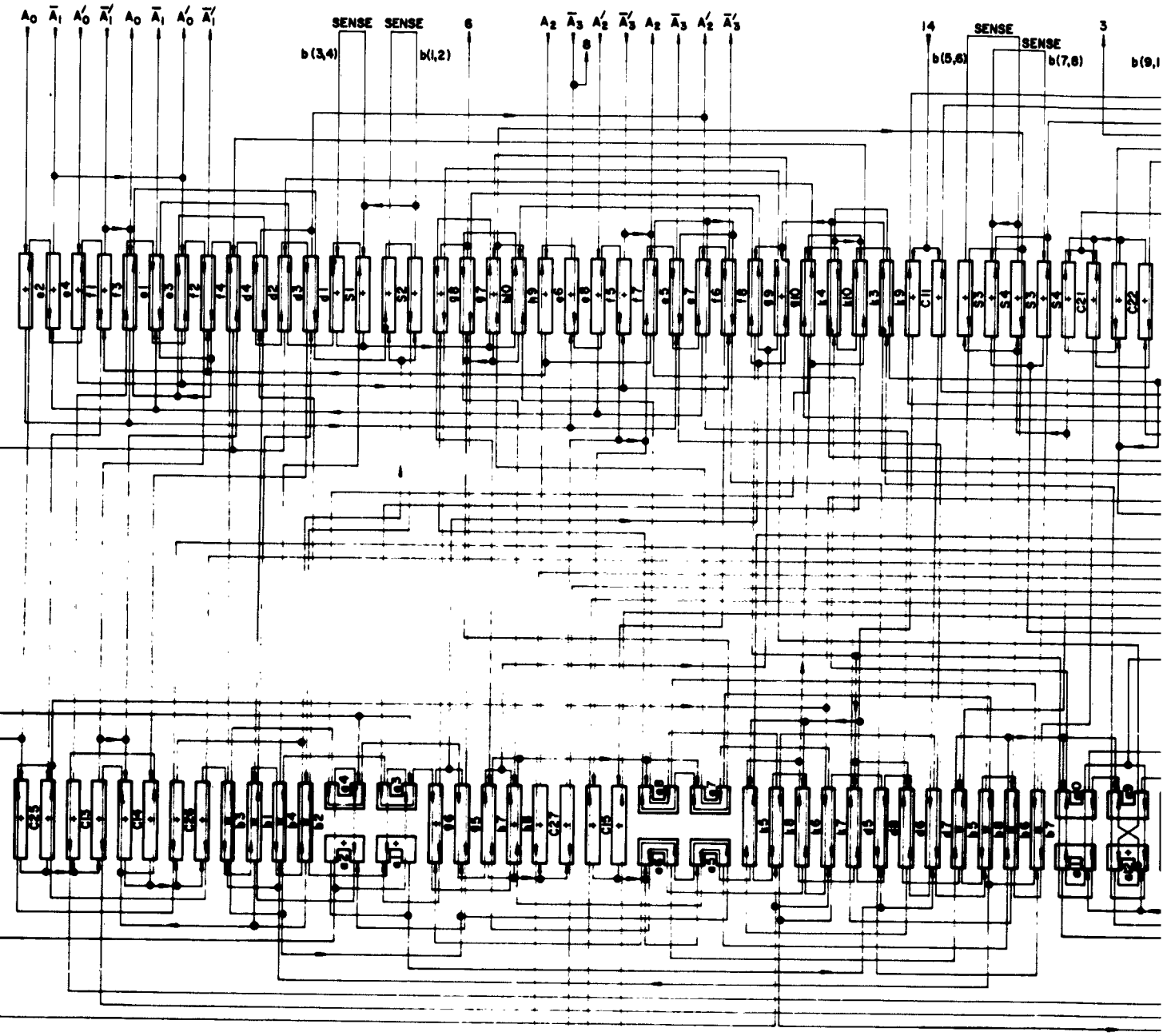
PHYSICAL LAYOUT OF THE SYSTEM

The integrated circuit layout for the memory control system is shown in Fig. 11a and 11b. The physical arrangement of the memory (Model B) and control system is shown drawn to scale in Fig. 12. The dimensions expressed in inches represent the size of the system if the gate and intergate wires are 0.006 inch wide, if the spaces between gates and between wires are 0.006 inch, and if the gate length-to-width ratio is 76. The dimensions are also expressed in units of gate widths w and gate lengths ℓ. If the gate width remains at 0.006 inch but the gate length-to-width ratio is reduced to 10, the height of the memory control system is reduced from 3.3 inches to 1.7 inches.

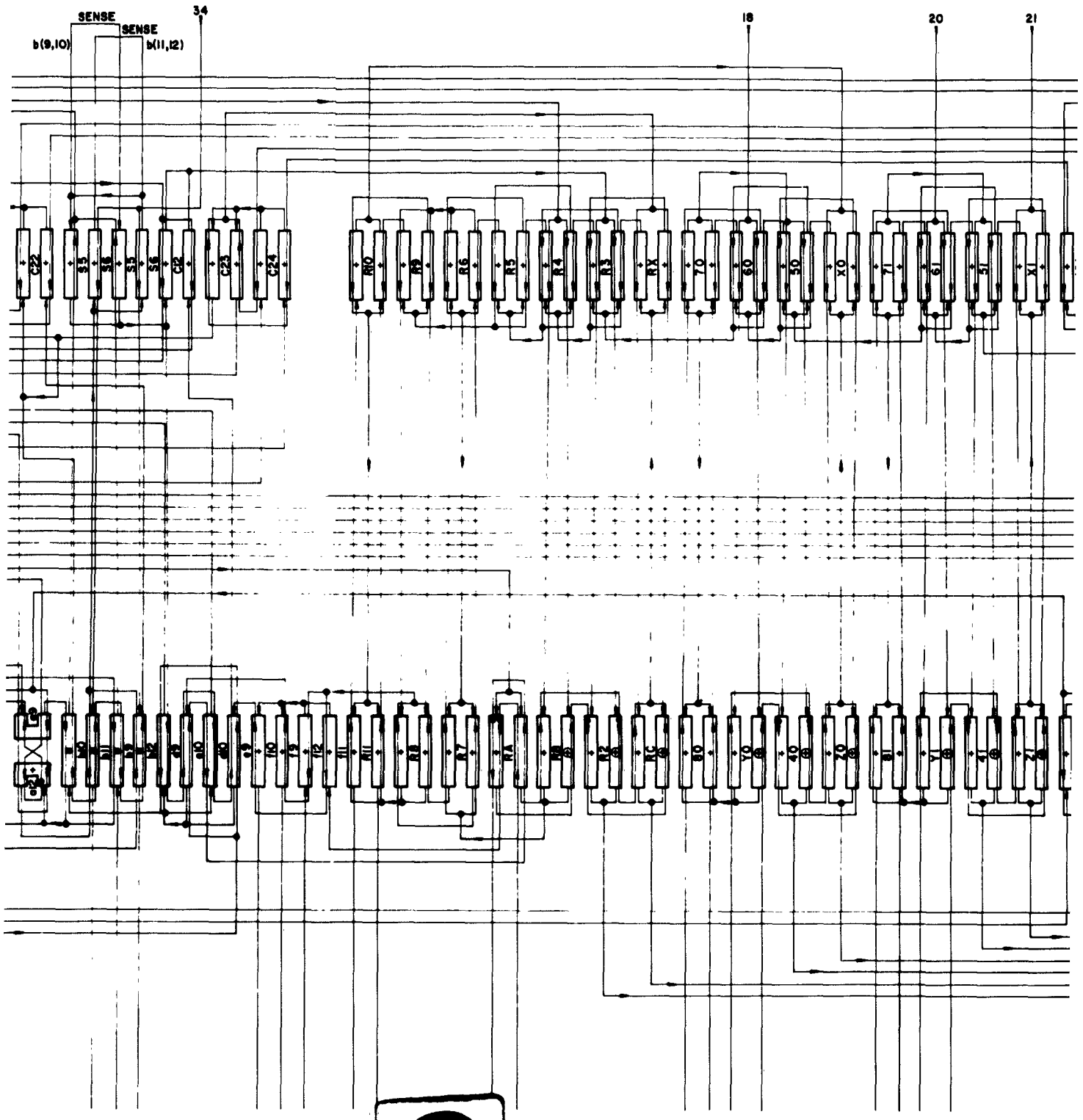
This memory control system contains 440 gates. Since all the circuit loops contain two gates, there are 220 loops. There are approximately 100 gates per square inch when the gate length-to-width ratio is 76.

SYSTEM OPERATION TIMES

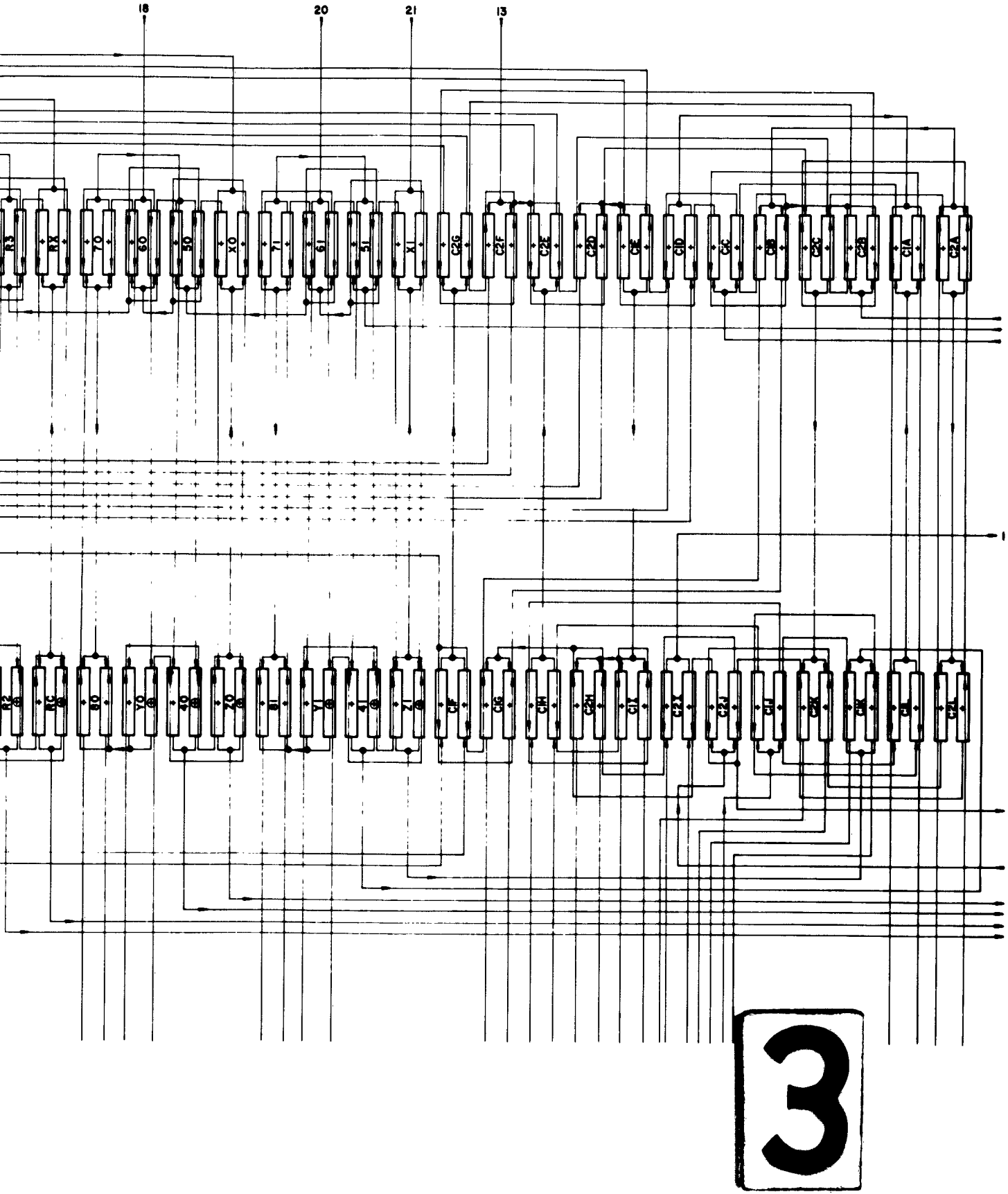
The circuit operating times are based on assumptions that were explained in detail in a previous report.[1] In essence, the inductance due to the gates and control lines, designated L_1, is



ure lla - Integrated circuit for the memory control system.



2


18
20
21
13
R3
RX
70
60
50
X0
71
61
51
X1
C2G
C2F
C2E
C2D
C2E
C1D
C4C
C3B
C2C
C2B
C1A
C2A
R2
RC
B0
Y0
40
Z0
B1
Y1
41
Z1
C1F
C1G
C1H
C2H
C1X
C2X
C2J
C1J
C2K
C1K
C1L
C2L
3

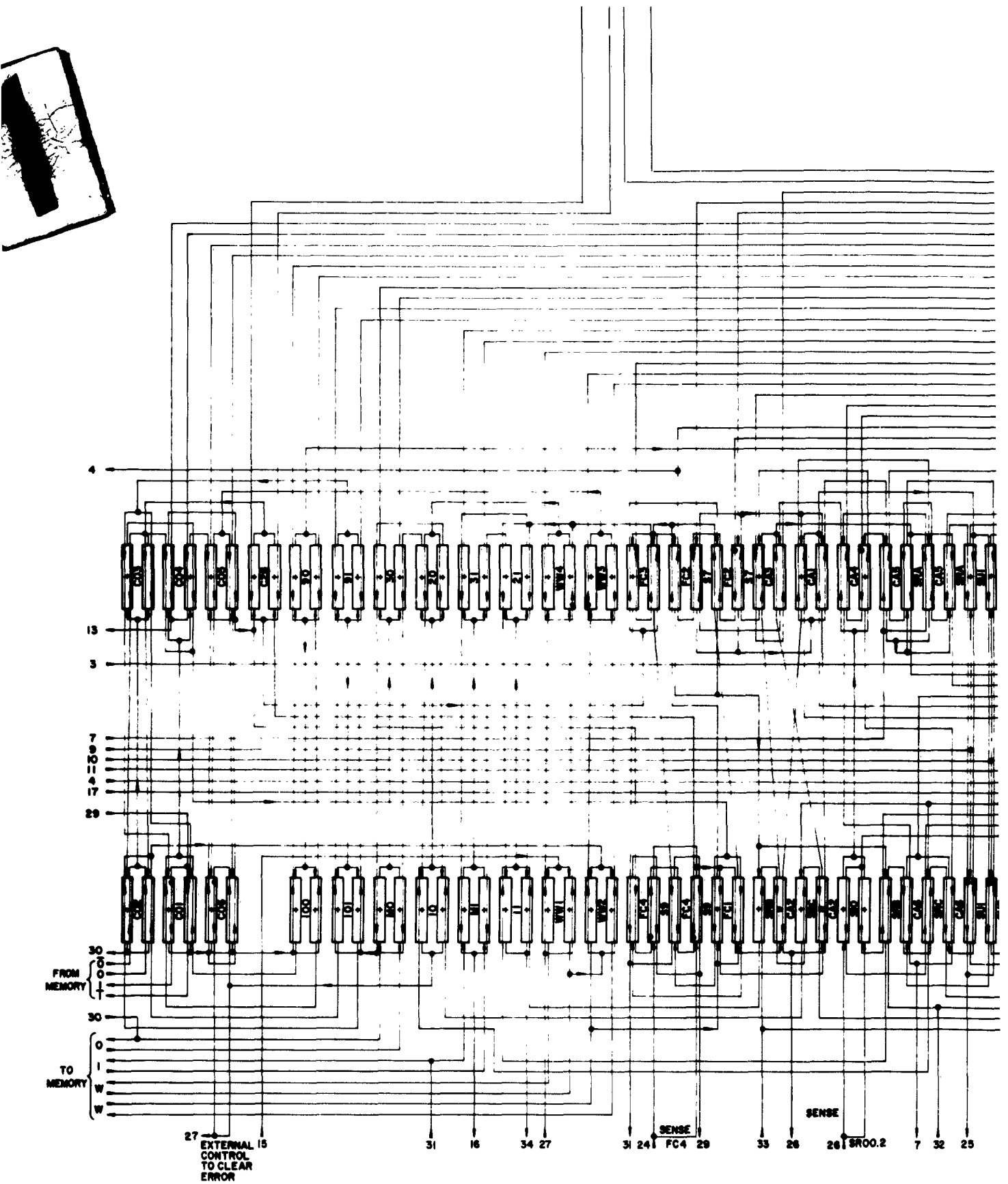
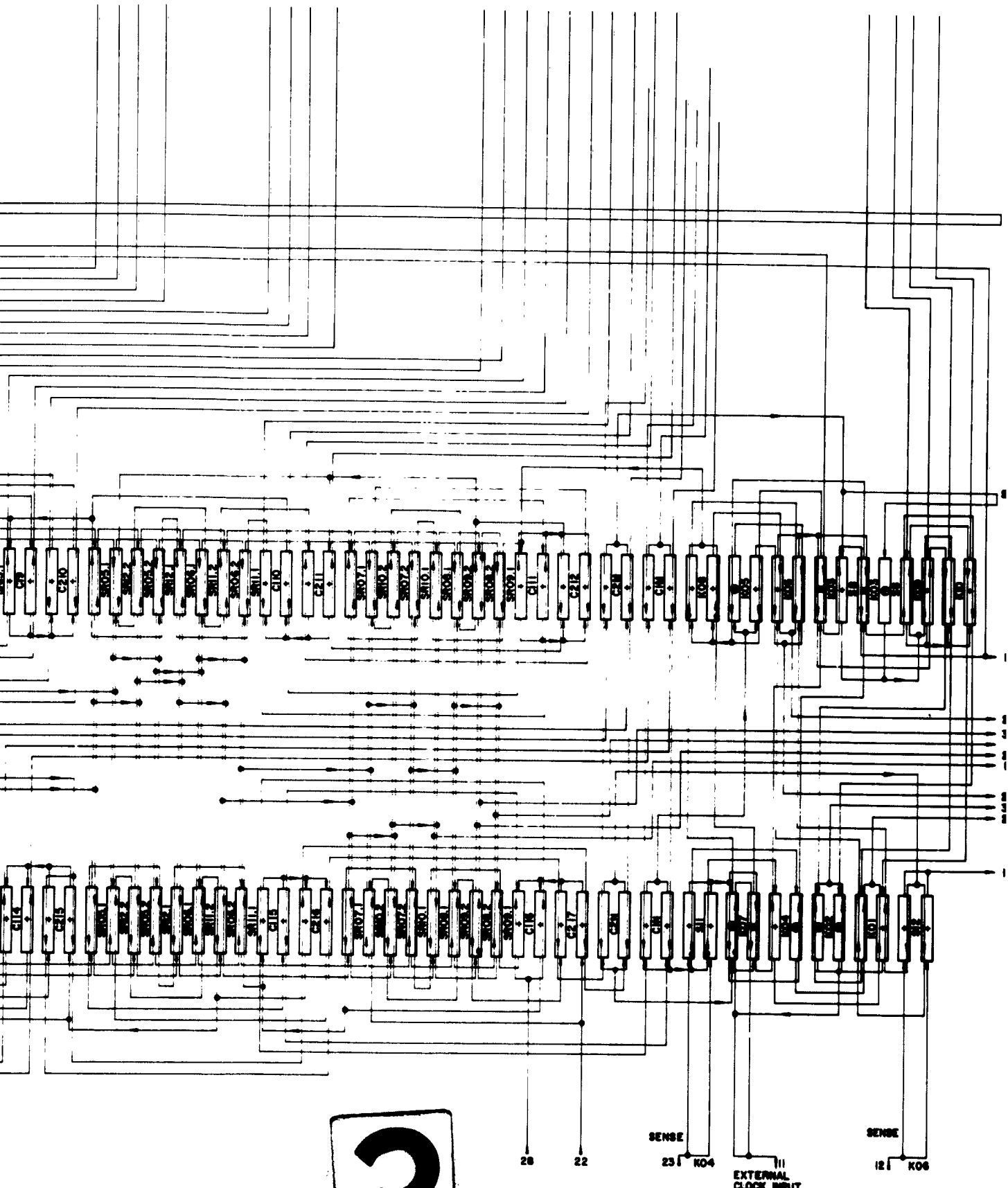


Figure 11b - Integrated circuit for the memory control system.

2


C20
C11
C12
C13
C14
C15
C16
C17
C18
C19
28
22
SENSE
23
KO4
11
EXTERNAL
CLOCK INPUT
SENSE
12
KO6

3

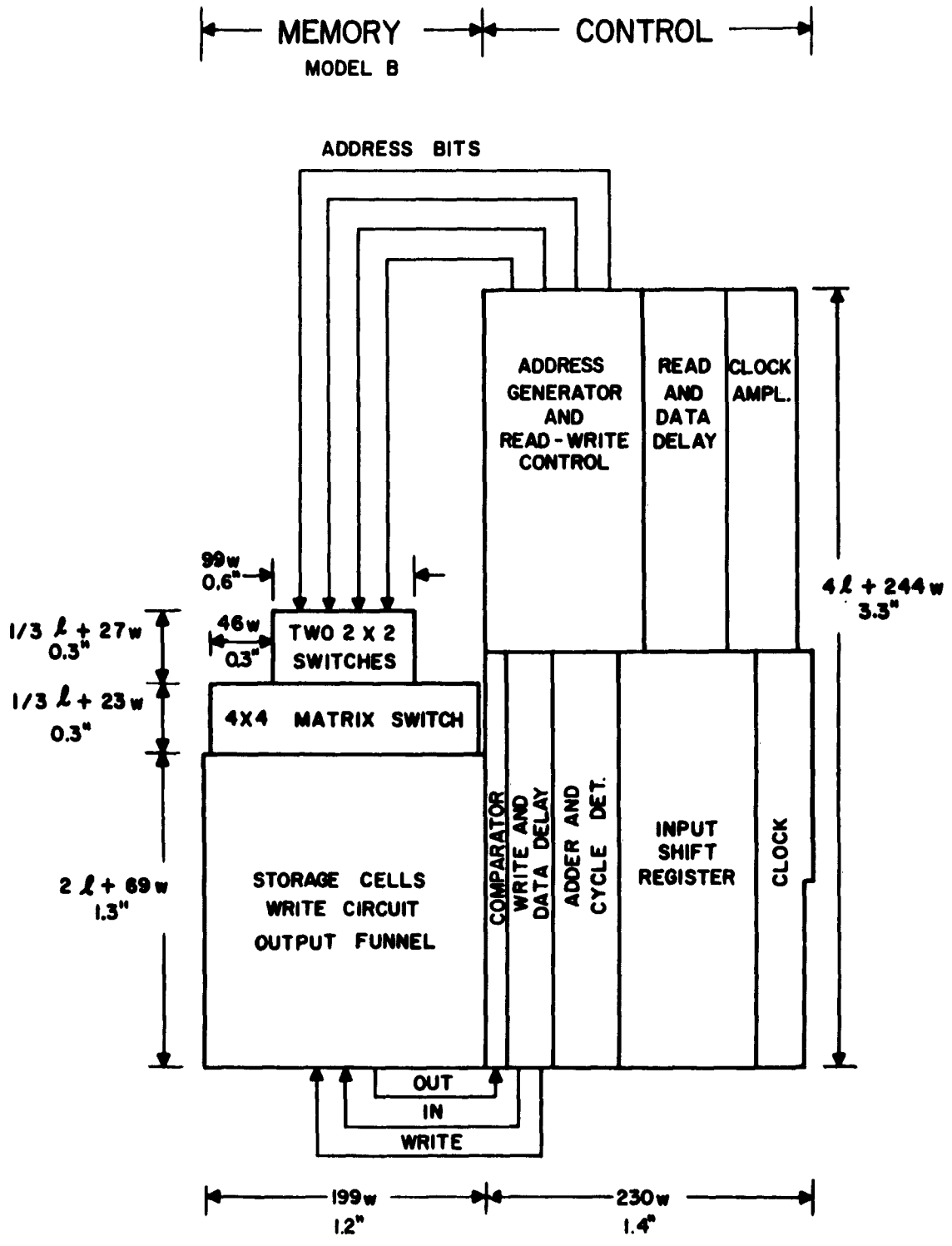


Figure 12 - Physical dimensions of the control system.

calculated for each important circuit loop. Also the inductance due to the intergate wiring, designated L_2, is calculated for each important circuit loop. The resistance of a normal gate is designated as R. The total inductance of a circuit is designated by L. The time constant of a circuit is calculated to be L/R equal to L_1/R plus L_2/R. The circuit operating time is taken to be two time constants (2L)/R.

Table I lists the statistics for the circuits of the control used with Model B of the memory that are used to determine the system operating times. The time constants indicated in the three right-hand columns of Table I are based on a gate length-to-width ratio ℓ/w of 10, a gate width w of 0.006 inch and a normal gate resistivity ρ of $(3)10^{-9}$ ohm meter. All the circuits listed in this table contain two gates. One of the two gates was assumed to be superconducting and the other resistive in the calculation of the inductance L_1.

Table II lists the operating times for the important parts of the memory control system. The information in this table was derived from the information in Table I. The operating times are listed for three cases. For a gate resistivity of $(3)10^{-9}$ ohm meter the operating times are based on gate length-to-width ratios of 10 and 76. It can be noted that the operating times of the circuits based on a gate length-to-width ratio of 10 are approximately double the

Table I - Circuit Statistics and Time Constants

Stage names	Number and condition[a] of controls: Lower control s	Lower control n	Upper control s	Upper control n	Number of units of intergate wiring N	Time constants in nanoseconds for $\ell/w = 10$ and $\rho = (3)10^{-9}$ ohm meter: L_1/R	L_2/R	L/R
K01	2	2	1	1	143	3.0	2.2	5.2
K02	1	2	2	1	171	3.2	2.6	5.8
K03	3	2	0	1	180	2.8	2.7	5.5
K04	2	1	0	1	82	2.1	1.2	3.3
K06	2	1	0	1	134	2.1	2.0	4.1
K08	1	0	0	1	43	1.3	0.6	1.9
K10	2	2	0	0	200	1.9	3.0	4.9
C2L	2	2	0	0	82	1.9	1.2	3.1
C2K	2	2	0	0	312	1.9	4.7	6.6
C2N	3	3	0	0	204	2.7	3.1	5.8
C2	2	2	0	0	90	1.9	1.4	3.3
$f(7,8)=\overline{A}'_3$	1	1	1	1	$(8\ell/3w)+643$	2.2	10.1	12.3
C01	1	0	0	1	108	1.3	1.6	2.9
C03	1	0	0	1	28	1.3	0.4	1.7
C04	1	0	0	1	28	1.3	0.4	1.7
C05	2	1	1	1	$(\ell/w)+629$	2.6	9.6	12.2
e(9, 10)	2	2	0	0	50	1.9	0.8	2.7
f(11, 12)	2	1	0	1	361	2.1	5.5	7.6
WW3	1	1	0	0	90	1.1	1.4	2.5
WW2	1	1	1	1	404	2.2	6.1	8.3
SRB	1	1	0	0	136	1.1	2.1	3.2
11	1	1	0	0	90	1.1	1.4	2.5
21	2	2	0	0	104	1.9	1.6	3.5
M1	1	1	1	1	404	2.2	6.1	8.3
31	3	3	0	0	472	2.7	7.1	9.8
Y1	1	0	0	1	126	1.3	1.9	3.2
61	1	1	0	0	20	1.1	0.3	1.4
71	1	1	0	0	126	1.1	1.9	3.0
81	1	1	0	0	459	1.1	6.9	8.0
91	1	1	0	0	90	1.1	1.4	2.5
101	1	0	0	1	108	1.3	1.6	2.9
X1	1	0	0	1	20	1.3	0.3	1.6

(continued)

Table I - (continued)

Stage names	Number and condition[a] of controls: Lower control		Upper control		Number of units of intergate wiring	Time constants in nanoseconds for $\ell/w = 10$ and $\rho = (3)10^{-9}$ ohm meter		
	s	n	s	n	N	L_1/R	L_2/R	L/R
51	1	0	0	1	28	1.3	0.4	1.7
RA	3	3	0	0	52	2.7	0.8	3.5
RB	1	0	0	1	126	1.3	1.9	3.2
R4	1	1	0	0	28	1.1	0.4	1.5
R5	1	1	0	0	20	1.1	0.3	1.4
R6	1	1	0	0	126	1.1	1.9	3.0
R7	1	1	0	0	20	1.1	0.3	1.4
R8	1	1	0	0	126	1.1	1.9	3.0
R9	1	1	0	0	20	1.1	0.3	1.4
R10	1	1	0	0	126	1.1	1.9	3.0
R11	1	0	0	1	402	1.3	6.1	7.4
RX	1	0	0	1	28	1.3	0.4	1.7
R3	1	0	0	1	24	1.3	0.4	1.7
b(9, 10)	3	3	2	2	497	4.9	7.5	12.4
d(1, 2)	2	1	1	2	183	3.2	2.8	6.0
a(1, 2)	2	1	1	2	273	3.2	4.1	7.3
g(9, 10)	1	1	1	1	194	2.2	2.9	5.1
a(11, 12)	1	0	0	1	72	1.3	1.1	2.4
SR01.1	1	0	0	1	$(2\ell/w)+172$	1.3	2.9	4.2
SU1	1	0	0	1	102	1.3	1.5	2.8
SU3	1	0	0	1	112	1.3	1.7	3.0
b(11, 12)	2	1	0	1	360	2.1	5.4	7.5
CA4	1	1	1	1	138	2.2	2.1	4.3
CA2	1	0	0	1	100	1.3	1.5	2.8
CA3	1	1	1	1	78	2.2	1.2	3.4
b(7, 8)	2	2	1	1	330	3.0	5.0	8.0
SR00.2	2	1	1	2	175	3.2	2.6	5.8

[a] s indicates a superconducting element, and n a normal (resistive) element. All circuits contain one normal and one superconducting gate.

Table II - Operating Times[a] for the Parts of the Control System

Gate resistivity ρ in ohm meters	$(3)10^{-9}$		$(12)10^{-9}$
Gate length-to-width ratio (ℓ/w)	10	76	76
Clock pulse generator (K01-K02-K03-K04)	39	24	6.0
Clock pulse generator, variation (K06-K08)	12	7.6	1.9
Clock amplifiers (K10-C2L-C2K-C2N-C2)	47	24	6.0
Address generator drivers ($f(7,8) = \overline{A}'_3$)	25	7.8	1.9
Memory word selector (4-pos. sel.) - (16-pos. sel.)	28	16	4.0
Memory output-funnel (1st through 5th level output)	33	16	4.0
Comparator part 1 (C01-C03)	9.2	5.8	1.4
Comparator part 2 (C04-C05)	28	11	2.7
Write delay to memory (e(9, 10)-f(11, 12)-WW3-WW2-W1)	58	26	6.5
Information delay to memory (SRB-11-21-M1-W1)	51	23	6.8
Information delay to comparator, minimum (SRB-11-21-31-Y1-61-71-81-91-101)	80	34	8.5
Information delay to comparator, variation (X1-51)	6.6	5.6	1.4
Read delay to comparator, minimum (e(9, 10)-f(11, 12)-RA-RB-R4-R5-R6-R7-R8-R9-R10-R11)	78	40	10
Read delay to comparator, variation (RX-R3)	6.8	5.4	1.4
Maximum C1 to C2 delay (d(1, 2)-a(1, 2)-g(9, 10)-a(11, 12))	42	22	5.5
Maximum C2 to C1 delay (b(9, 10))	25	12	3.0
Maximum shift-register delay (SR01.1-SU1-SU3)	20	9.4	2.4
Carry delay from counter to adder (b(11, 12)-CA4)	24	11	2.7
Adder carry delay (CA2-CA3-CA4)	21	14	3.5

[a]The circuit operating time in nanoseconds is assumed to be equal to two circuit time constants (2L/R).

operating times of the circuits based on the longer gates. It should be pointed out that the circuits were arranged on Fig. 11 so that a minimum number of intergate wires traveled the length of a gate. This arrangement favored the circuit operating time for the longer gates. Table II also lists circuit operating times based on a gate resistivity of $(12)10^{-9}$ ohm meter and a gate length-to-width ratio of 76. These circuit operating times are simply one quarter of those when the gate resistivity is $(3)10^{-9}$ ohm meter and the gate length-to-width ratio is 76.

Figure 13 shows a timing diagram for the entire system that is based on information in the middle column of operating times in Table II. The length of each box on this diagram is proportional to the time delay through the box. The height of the boxes has no significance. Figure 13 intends to show the timing relationship of the various parts of the system. The delay indicated by each box represents the maximum fixed delay through the box. Certain boxes have a dotted end on them. This indicates additional delay that can be added to the box under external control. This controlled variable delay is available in order to more exactly match the delays of various chains in the system to one another.

During writing, clock signals originate in the clock pulse generator and are fanned out through the clock amplifiers to final

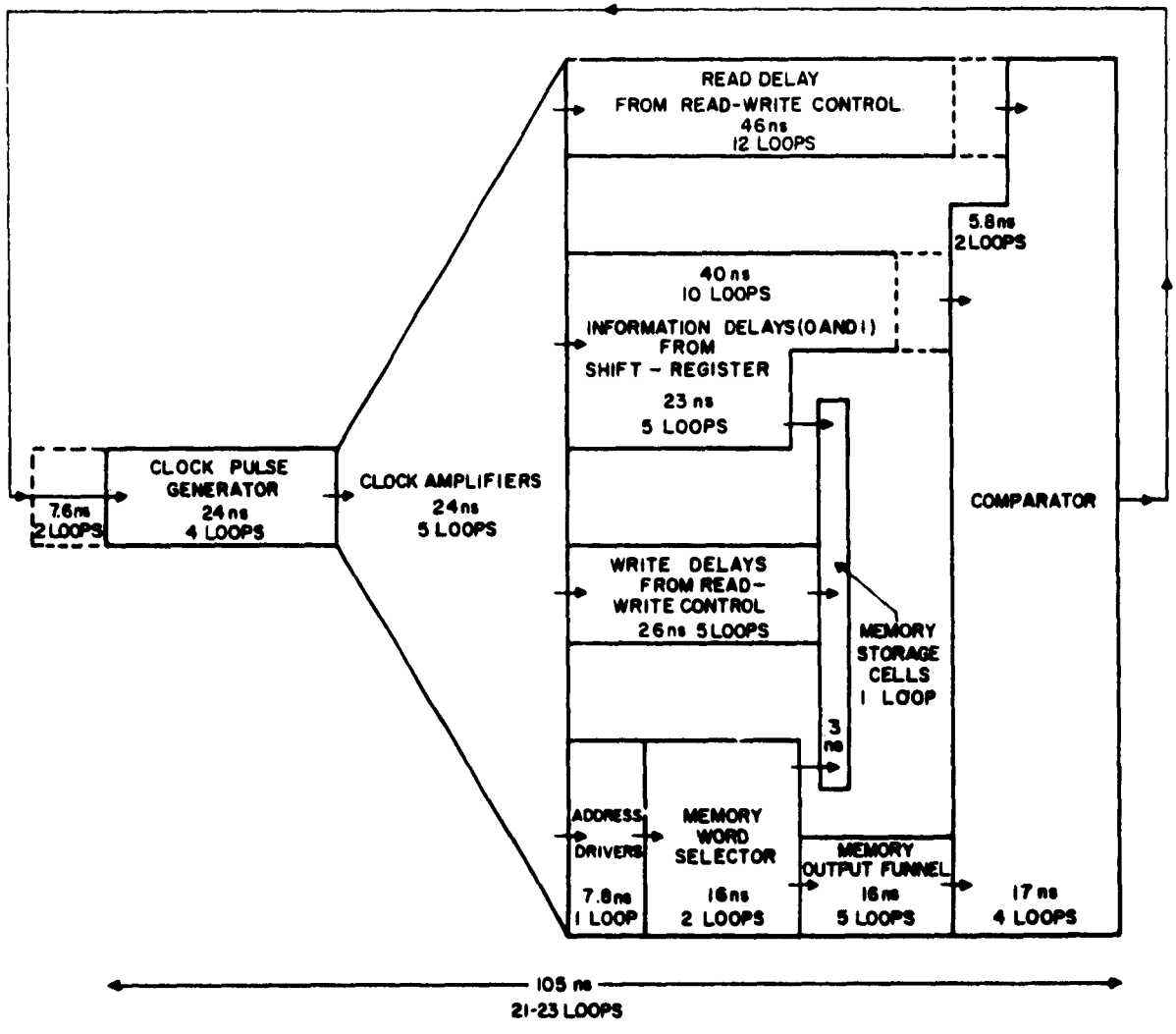


Figure 13 - Timing diagram for the control system.

clock pulse drivers. There are 16 final C1 clock driver circuits and 17 final C2 clock driver circuits. These final clock driver circuits are intended to simultaneously initiate signals in five parallel chains that have a common destination in a memory storage cell. These five chains are an address driver and memory word selector chain, an information "0" chain from the shift-register, an information "1" chain from the shift-register, and a pair of write signal chains from the read-write control flip-flop. On Fig. 13 the two information chains are combined in one box, and the two write chains are combined in a second box. The delays for the case shown in this figure are 23 nanoseconds for the shortest chain and 26 nanoseconds for the longest chain.

During reading, signals again originate in the clock pulse generator, fan out through the clock pulse amplifier, and are distributed to four chains. One chain includes the path from the address driver, through the memory word selector and output funnel, to the input, to the comparator. A pair of chains intended to match the time delay of the first chain originates in the shift-register, passes through delay stages, and terminates as an input to the comparator. There is one chain for "0" and a second chain for "1." These two chains can be varied from a minimum of 34 up to a maximum of 40 nanoseconds under external control, and thus can be made to match the 40-nanosecond

path through the memory. The memory and information chains join in the comparator and pass through two circuit loops in the comparator before they join with a fourth chain. The fourth chain is a read-signal delay chain that originates in the read-write flip-flop and terminates in the comparator. This read chain can be varied from a minimum of 40 up to a maximum of 47 nanoseconds and thus can be made to match the delay of 46 nanoseconds for the chain from the address drivers through the first half of the comparator. The signal from the output of the comparator feeds back into the input to the clock pulse generator. The rate of the clock pulse generator is not limited by the path through the comparator since it cycles through its own self-contained ring. If the clock is running at its maximum rate of one pulse every 24 nanoseconds, an error detected by the comparator would stop the clock on the fourth clock signal after the clock signal which had originated the detection of the error.

The maximum clock pulse rate is determined by the maximum delay of signals originating with one clock phase and terminating in circuits clocked with the other phase. Table II lists a number of maximum circuit paths for various subunits of the system. The longest delay of the type that determines the clock pulse rate occurs in the carry circuit path of the address generator. This path

originates with circuit d(1, 2) and terminates with circuit a(11, 12). The time delay of this path is 22 nanoseconds for $\rho = (3)10^{-9}$ and $\ell/w = 76$. The switching of the clock driver signals does not occur simultaneously since the paths through the clock amplifier are not quite equal. An analysis of these clock amplifier chains indicated a time-delay variation from a minimum of 22 up to a maximum of 24 nanoseconds. In order to allow circuits controlled, directly or indirectly, by one phase to settle to their proper state before the other phase occurs, the 2-nanosecond clock-driver delay variation is added to the 22-nanosecond maximum circuit path of the carry for the address generator to give a 24-nanosecond minimum timing between clock pulses. The minimum circuit path for the clock ring is 24 nanoseconds. The clock ring delay can be varied under external control up to 32 nanoseconds. Still slower clock rates can be obtained by bringing a clock trigger pulse in from an external source.

The variation in delays among parts of the system intended to have equal delays is considerably worse for $\ell/w = 10$ and the same gate resistivity. In this case, the clock amplifier path delays varied from a minimum of 42 up to 55 nanoseconds for the same wire routes. The variation of the delay among the clock amplifier paths increased from a $\pm 4\%$ variation for $\ell/w = 76$ up to $\pm 13\%$ variation for $\ell/w = 10$. This increase in time variation reflects the increase in the proportion

of intergate wiring inductance (L_2) relative to the total circuit inductance L as the length of the gate decreases.

TableIII summarizes the operating times for the entire system and certain subunits of the system. Such subunits as the address generator and shift-register can be operated together with a clock pulse generator independent of the rest of the system. The operating times for these subunits, indicated in TableIII, reflect this assumption.

OPTIMUM FAN-OUT

Introduction

In most instances the detailed logic of a system can be designed in a number of ways. Where system speed is of primary importance, the optimum design may hinge upon the optimum choice of fan-out. The succeeding paragraphs discuss the various factors affecting the optimum fan-out and present pertinent calculations for a number of representative cases. The discussion also considers the effect of deviating from the optimum fan-out upon the speed of a set of circuits and upon the number of circuits.

Table III - Summary Operating Times[a] for the Memory and Control System

Gate resistivity ρ in ohm meters	$(3)10^{-9}$		$(12)10^{-9}$
Gate length-to-width ratio (ℓ / w)	10	76	76
Memory write delay (4-pos. sel.)-(16-pos. sel.)-(storage)	32	19	4.7
Memory read delay (4-pos. sel.)-(16-pos. sel.)-(output funnel)	61	32	8.0
System write delay (clock gen.)-(clock amplifier)-(write delay from read-write control)	148	77	19
System read delay (clock gen.)-(clock amplifier)-(address drivers)-(4-pos. sel.)-(16-pos. sel.)-(output funnel)	172	88	22
System read and check delay (clock gen.)-(clock amp.)-(address drivers)-(4-pos. sel.)-(16-pos. sel.)-(output funnel)-(comparator)	209	105	27
Minimum clock period (C1+C2)	78	48	12
Maximum address generation period[b] for address generator (d(1, 2)-a(1, 2)-g(7, 8)-a(7, 8)-b(7, 8))	57	28	7.0
Maximum shifting period[b] for shift-register (SR01. 1-SR00. 2)	20	11	2.7
Maximum shifting and adding period[b] for shift-register and adder (SR01. 1-SU1-SU3)+(CA2-CA3-CA4)	41	25	6.2

[a] The operating times expressed in nanoseconds are based on the sums of the operating times of individual circuits listed in Tables I and II.

[b] The indicated operating times would apply if these subunits were operated independently of the rest of the system.

The calculations that follow refer specifically to in-line cryotron circuits of the type described in the ninth quarterly Lightning progress report.[1] They are, however, in a general sense applicable to other types of circuits as well.

The in-line cryotron referred to above has a two-path loop consisting of:

1. one or more gates of length ℓ in each path, each gate being controlled by one or two signals plus the necessary dc bias,
2. one or more controls (destinations of the loop signals) of length ℓ,
3. loop wiring in length units w (width of line).

The values of the various types of delay that may be associated with a cryotron loop are compiled in Table IV. They are based on the operation time calculations on pp. 217-227 of the Lightning report.[1]

Delay of a Cryotron Loop

Basically, the delay of a cryotron loop D_{loop} can be divided into three categories:

1. a constant delay A independent of fan-out and consisting of gates or controls,

Table IV - Delays of In-Line Cryotron Loop

Type of Delay	Symbol	Delay (nanoseconds)
Normal gate of length ℓ	D_{ng}	0.38
Superconducting gate of length ℓ	D_{sg}	0.35
Upper control over a normal gate	D_{un}	1.16
Upper control over a superconducting gate	D_{us}	1.01
Lower control over a normal gate	$D_{\ell n}$	0.85
Lower control over a superconducting gate	$D_{\ell s}$	0.70
Length w of wiring	D_w	$0.30 \frac{w}{\ell}$

2. a delay BF directly proportional to the fan-out and consisting of gates or controls where F is the fan-out and B a constant,

3. wiring delay.

The last category needs elaboration. The amount of delay contributed by the wiring can be determined only by actually laying out systems of cryotron circuits. The following factors, however, influence the wiring length of a loop.

1. The wiring length of a loop is an increasing function of the number of gates and controls in the loop and, hence, an increasing function of the fan-out. The function, however, is less than linear when wiring paths that connect a signal to its destination are selected by means of a wire length minimization scheme. For example, to connect a signal s from a source to only one destination d_1, the distance sd_1 is used. However, to connect a source to two destinations, d_1 and d_2, the minimum of the two paths, sd_1d_2 and sd_2d_1, can be selected. With more destinations per signal, a greater choice of paths and hence a smaller wire length contribution per connection is possible.

2. The wiring length of a loop is an increasing function of the number of gates in the entire system of loops. The number of gates in the system in turn is a decreasing function of the fan-out. Moreover, if the fan-out is increased the number of loops decreases,

the number of gates and controls per loop increases, and the wiring length per gate and control decreases as pointed out in 1. The wiring length is, therefore, a decreasing function of the fan-out.

As a result of these considerations, the assumption is made that the total number N_w of wiring sections of length w is independent of the fan-out. The wiring delay of a loop in nanoseconds becomes

$$D_{wiring} = N_w D_w = 0.30 N_w \frac{w}{\ell}. \quad (1)$$

The loop delay then becomes

$$D_{loop} = A + BF + 0.30 N_w \frac{w}{\ell}. \quad (2)$$

Equations for Optimum Fan-Out

The optimum fan-out is computed by minimizing the total delay D_{total} of a pertinent signal through a series of cryotron loops. The total delay is, of course, the sum of individual loop delays in the series. In general, the various loops may perform different logical functions, and thus their optimum fan-outs (and hence their delays) may differ from each other. As a first step, however, all loops in the series will be considered as performing the same logical

function. If L is the number of loops in the series

$$D_{total} = LD_{loop} = L \left(A + BF + 0.30N_w \frac{w}{l} \right). \qquad (3)$$

For optimum fan-out

$$\frac{d\, D_{total}}{d\, F} = D'_{total} = LB + \left(A + BF + 0.30N_w \frac{w}{l} \right) L' = 0, \qquad (4)$$

so that

$$- \left[\frac{L}{L'} + F \right]_{opt} = \frac{A + 0.30N_w \frac{w}{l}}{B} \qquad (5)$$

where the subscript opt refers to the values at optimum fan-out.

Effect of Using Larger-Than-Optimum Fan-Out

In some cases, it may be desirable to sacrifice some speed in a set of cryotron loops in order to reduce the number of cryotrons in the set. For example, a system consisting of a number of substrates of specified size contains inherent delays due to such substrate connections. Using a larger-than-optimum fan-out in various sub-systems may reduce the number of substrates. As a result, the higher system speed resulting from the reduction in the

number of substrate connections may more than offset the lower sub-system speeds resulting from the increased fan-out.

To show the effect of increased fan-out on the delay through a set of cryotron loops, the percent increase of delay at some multiple of optimum fan-out over the delay at optimum fan-out is computed for various multiples of optimum fan-out for the cases presented below. At the same time, the corresponding percent change in the required number of cryotrons is also computed.

Case I: Repeater

A single loop is to be replicated R times through a series of repeater levels as illustrated in Fig. 14 and 15. At each level, each loop fans out F times. Therefore,

$$R = F^{L}, \tag{6}$$

$$L = \frac{\ln R}{\ln F}, \tag{7}$$

$$\frac{dL}{dF} = L' = \frac{-\ln R}{F(\ln F)^2} = -\frac{L}{F \ln F}, \tag{8}$$

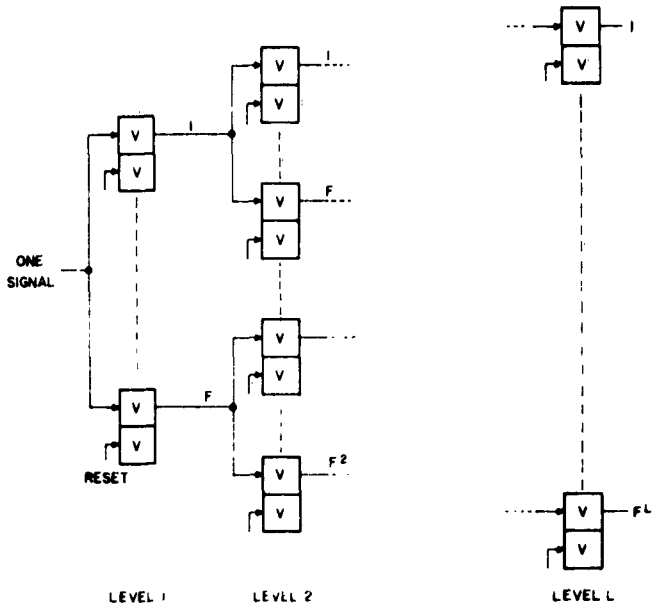


Figure 14 - Repeater signals with separate resetting.

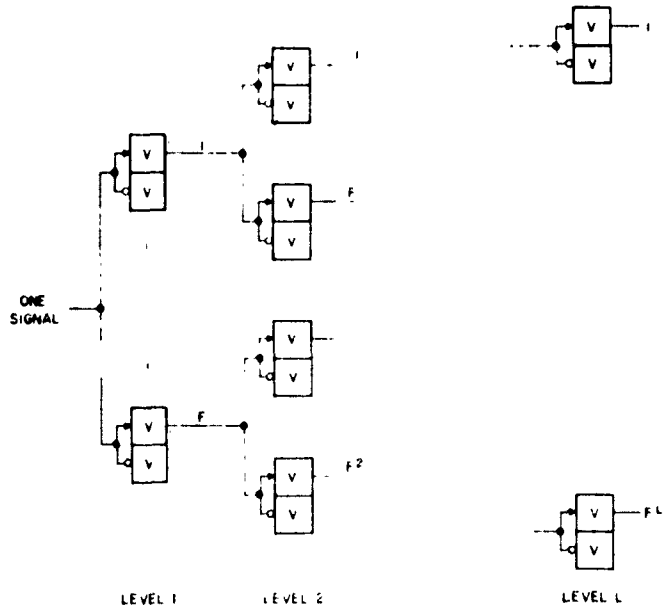


Figure 15 - Repeater signals with self-resetting.

and

$$-\left[\frac{L}{L'} + F\right]_{opt} = F_{opt}(\ln F_{opt} - 1). \qquad (9)$$

From Eqs. (5) and (9):

$$F_{opt}(\ln F_{opt} - 1) = \frac{A + 0.30N_w \frac{w}{l}}{B} = C. \qquad (10)$$

If the fan-out used is k times the optimum fan-out, the percent increase in total delay Δ_{delay} becomes by means of Eqs. (3), (7) and (10):

$$\Delta_{delay} = 100\left[\frac{D_{total}}{D_{total\ opt}} - 1\right] = 100\left[\frac{\frac{\ln R}{\ln F}(BC + BF)}{\frac{\ln R}{\ln F_{opt}}(BC + BF_{opt})} - 1\right]$$

$$= 100\left[\frac{(C+kF_{opt})\ln F_{opt}}{(C+F_{opt})\ln(kF_{opt})} - 1\right] = 100\left[\frac{(C+F_{opt})\ln F_{opt} + (k-1)(C+F_{opt})}{(C+F_{opt})\ln(kF_{opt})} - 1\right]$$

$$= 100\left[\frac{\ln F_{opt} + k - 1}{\ln(k F_{opt})} - 1\right] = 100\left[\frac{k - \ln k - 1}{\ln(k F_{opt})}\right]. \qquad (11)$$

Similarly, the ratio of the number of levels becomes:

$$\frac{L}{L_{opt}} = \frac{\ln F_{opt}}{\ln (k F_{opt})}, \tag{12}$$

and the percent change in the number of cryotrons required Δ_{cr} becomes:

$$\Delta_{cr} = 100\left[\frac{N_{cr}}{N_{cr\,opt}} - 1\right] = 100\left[\frac{2(F^1 + F^2 + \ldots + F^L)}{2(F_{opt}^1 + F_{opt}^2 + \ldots + F_{opt}^{L_{opt}})} - 1\right]$$

$$= 100\left[\frac{\frac{F(F^L - 1)}{(F - 1)}}{\frac{F_{opt}(F_{opt}^{L_{opt}} - 1)}{(F_{opt} - 1)}} - 1\right] = 100\left[\frac{F}{F_{opt}}\,\frac{(F_{opt} - 1)}{(F - 1)}\,\frac{(F^L - 1)}{(F_{opt}^{L_{opt}} - 1)} - 1\right]$$

$$= 100\left[\frac{k(F_{opt} - 1)}{(kF_{opt} - 1)} - 1\right] = 100\left[\frac{(1 - k)}{(kF_{opt} - 1)}\right], \tag{13}$$

where N_{cr} and $N_{cr\,opt}$ are the number of cryotrons required with non-optimum and with optimum fan-out respectively.

a. As a first approximation, let D_{loop} be equal to F units of delay so that $A = N_w = 0$. Then from Eq. (10)

$$\ln F_{opt} = 1, \tag{14}$$

$$F_{opt} = \epsilon = 2.7. \tag{15}$$

b. Let each loop of Fig. 14 consist of:

1. one normal and one superconducting gate ($A = D_{ng} + D_{sg} = 0.73$),
2. F lower controls over normal gates ($B = D_{ln} = 0.85$),
3. wiring delay.

Resetting of the loop is assumed to be done by a separate resetting signal as described in the Proposal for Phase IV of Lightning.[2]

From Eq. (10)

$$F_{opt} (\ln F_{opt} - 1) = 0.85 + 0.35 N_w \frac{w}{l}. \tag{16}$$

c. Let each loop of Fig. 15 consist of:

1. one normal and one superconducting gate ($A = D_{ng} + D_{sg} = 0.73$),
2. F lower controls over superconducting gates and F lower controls over normal gates ($B = D_{ln} + D_{ls} = 1.55$),
3. wiring delay.

Resetting of the loop is assumed to be done by the setting signal acting as an input dual (control current is in the same direction as the reset gate current).

From Eq. (10):

$$F_{opt} (\ln F_{opt} - 1) = 0.47 + 0.193 N_w \frac{w}{\ell} \tag{17}$$

Case II: Funnel

A set of R mutually exclusive loop signals is to be funneled into one loop through a series of levels as illustrated in Fig. 16 and 17. At each level F loop signals are OR-gated into one loop of the succeeding level. The number of levels is, therefore, L. Eqs. (6) through (12) are applicable.

a. Let each loop of Fig. 16 consist of:

1. one normal gate and F superconducting gates for resetting one gate and for OR-gating F input signals ($B = D_{sg} = 0.35$),
2. one lower control over a normal gate (the slowest loop) of the succeeding level ($A = D_{ng} + D_{\ell n} = 1.23$),
3. wiring delay.

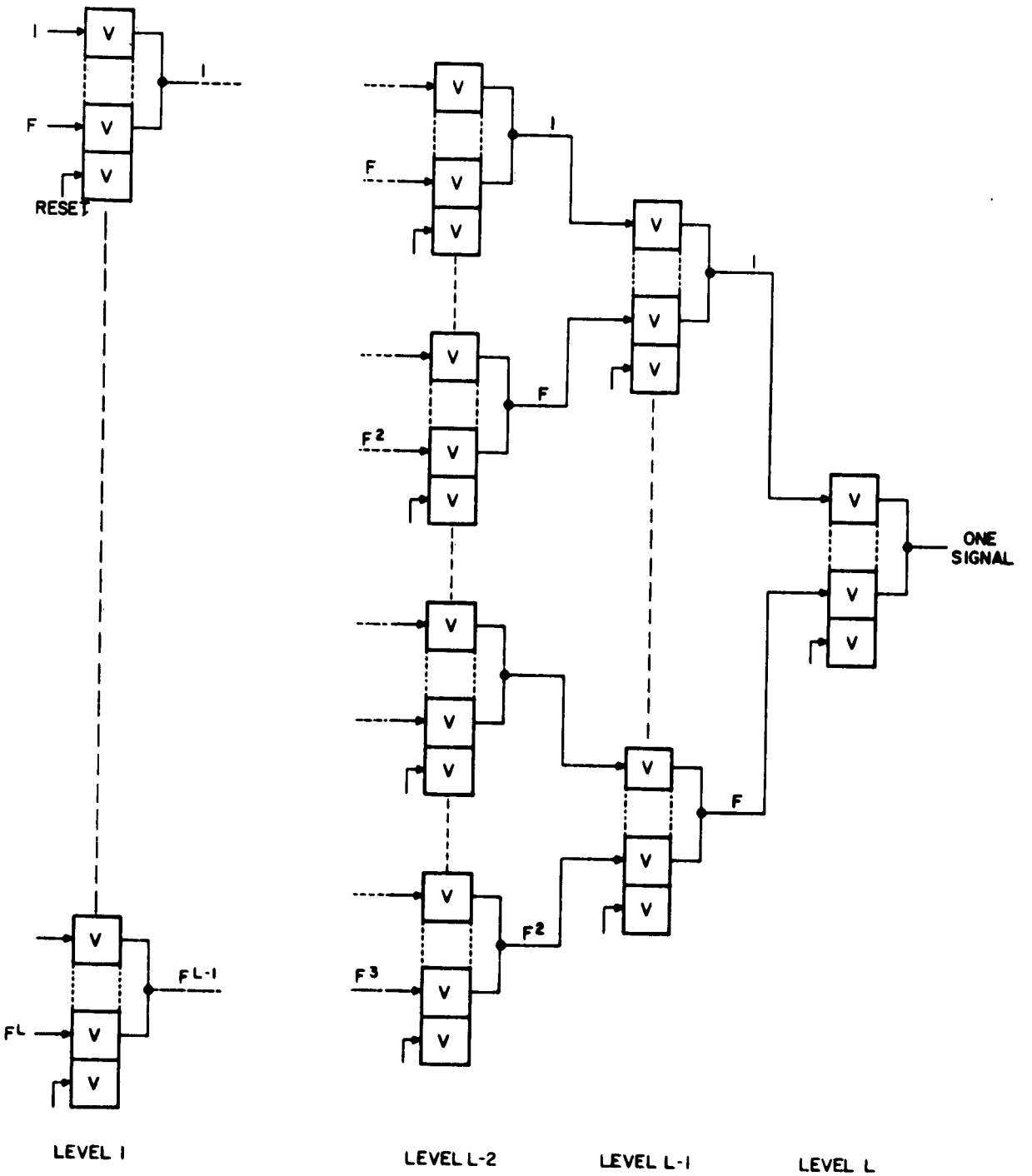


Figure 16 - Funnel with one-input gates.

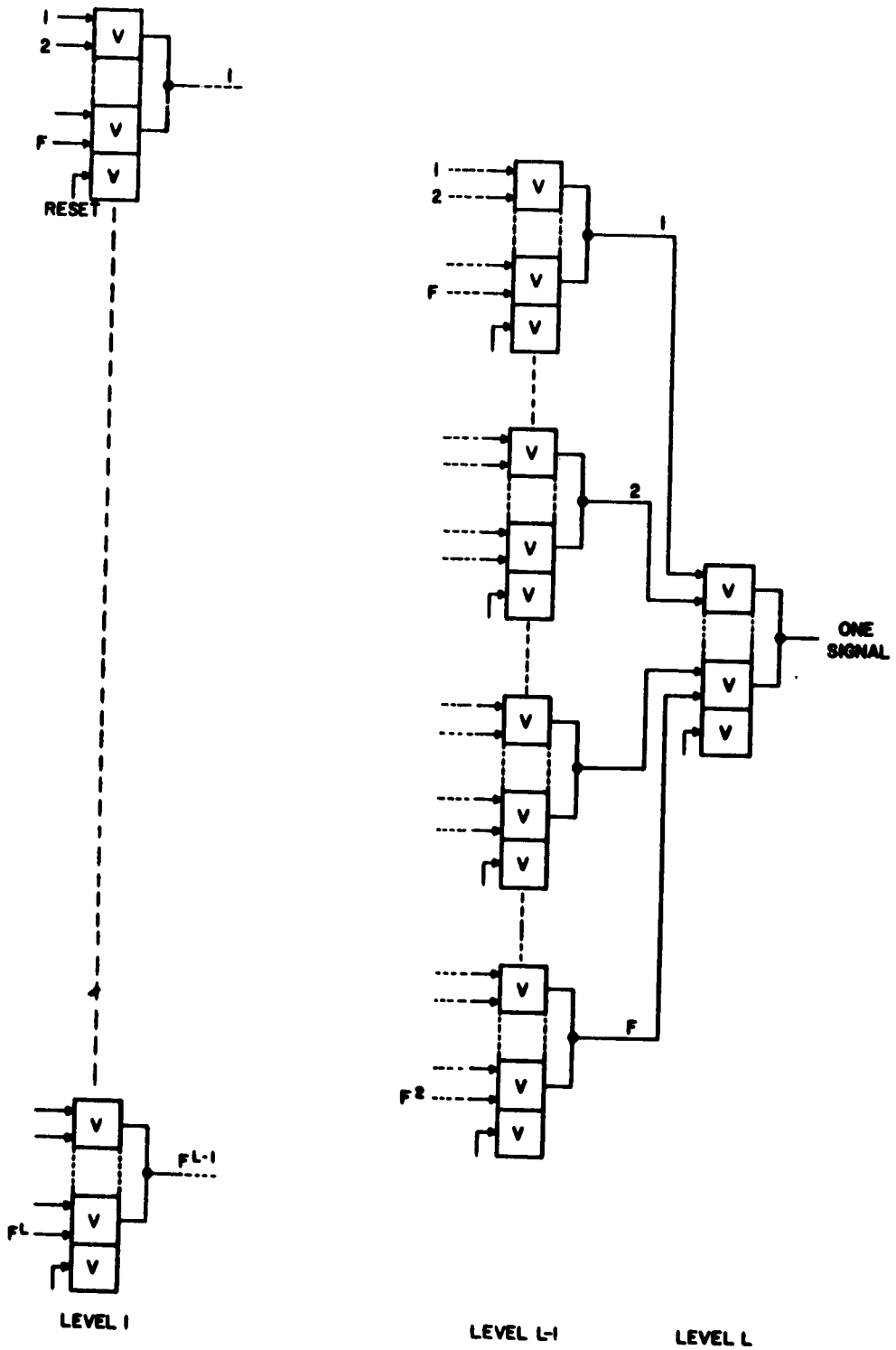


Figure 17 - Funnel with two-input gates.

Resetting of the loop is assumed to be done by a separate resetting signal. As before:

$$F_{opt}(\ln F_{opt} - 1) = 3.5 + 0.86N_w \frac{w}{l}, \tag{18}$$

and

$$\Delta_{cr} = 100\left[\frac{N_{cr}}{N_{cr\ opt}} - 1\right] = 100\left[\frac{(F+1)(F^{L-1}+\ldots+F+1)}{(F_{opt}+1)(F_{opt}^{L_{opt}-1}+\ldots+F_{opt}+1)} - 1\right]$$

$$= 100\left[\frac{2(1-k)}{k(F_{opt}+1) - 1 - \frac{1}{F_{opt}}}\right]. \tag{19}$$

b. Let each loop of Fig. 17 consist of:

1. one normal gate and F/2 superconducting gates for resetting one gate and for OR-gating F input signals by means of F/2 two-input gates ($B = \frac{D_{sg}}{2} = 0.176$),
2. one upper control over a normal gate (the slowest loop) of the succeeding level ($A = D_{ng} + D_{un} = 1.54$),
3. wiring delay.

Resetting of the loop is assumed to be done by a separate resetting signal. Then

$$F_{opt}\,(\ln F_{opt} - 1) = 8.7 + 1.71 N_w \frac{w}{\ell}, \tag{20}$$

and

$$\Delta_{cr} = 100\left[\frac{N_{cr}}{N_{cr\ opt}} - 1\right] = 100\left[\frac{\left(\frac{F}{2}+1\right)(F^{L-1}+\ldots+F+1)}{\left(\frac{F_{opt}}{2}+1\right)(F_{opt}^{L_{opt}-1}+\ldots+F_{opt}+1)} - 1\right]$$

$$= 100\left[\frac{3(1-k)}{k(F_{opt}+2) - 1 - \frac{2}{F_{opt}}}\right]. \tag{21}$$

Numerical Results and Conclusions

Table V presents the values of fan-out for the various cases discussed. They indicate that the wiring delay, as assumed, exerts considerable influence on the optimum fan-out. For example, in the case of the repeater with separate reset signal (Ib), a length-to-width ratio ℓ/w of 10 and a loop wiring length of 100w increase the optimum fan-out from the no-wiring-delay value of 3.5 to 5.8.

Table V - Optimum Fan-Out

Case \ N_w $\frac{w}{l}$	0	10	20	30	40	50
Ia	2.7	-	-	-	-	-
Ib	3.5	5.8	7.7	9.3	10.8	12.3
Ic	3.2	4.6	5.8	6.8	7.8	8.7
IIa	5.3	9.6	13.1	16.3	19.3	22.0
IIb	8.0	15.0	21.0	26.0	32.0	36.0

TableVI shows quantitatively the effect of using a larger-than-optimum fan-out on both the total delay and the number of cryotrons. Two significant facts emerge: 1) the total delay increases relatively slowly with increased fan-out; and 2) one can increase the actual fan-out by a considerable factor before the increase in delay fails to be compensated by a decrease in the number of cryotrons. The latter consideration is especially applicable to the funnel. Coupled with the fact that the funnel has the larger optimum fan-out to begin with, a considerably larger fan-out for the funnel than for the repeater is warranted.

Table VI - Effect of Larger-Than-Optimum Fan-Out on Delay and on Number of Cryotrons

F_{opt}	$k = \frac{F}{F_{opt}}$	Δ_{delay} in %	Δ_{cr} in %		
			Cases Ib and Ic	Case IIa	Case IIb
2	2	+ 22	-33	-44	-50
	4	+ 78	-43	-57	-64
	8	+178	-47	-62	-70
	16	+352	-48	-65	-73
4	2	+ 15	-14	-23	-29
	4	+ 58	-20	-32	-40
	8	+142	-23	-36	-45
	16	+294	-24	-38	-48
6	2	+ 12	- 9	-16	-20
	4	+ 50	-13	-23	-29
	8	+127	-15	-26	-34
	16	+267	-16	-27	-36
8	2	+ 11	- 7	-12	-16
	4	+ 47	-10	-17	-23
	8	+118	-11	-20	-27
	16	+252	-12	-21	-28
10	2	+ 10	- 5	-10	-13
	4	+ 43	- 8	-14	-19
	8	+113	- 9	-16	-22
	16	+241	- 9	-17	-24

REFERENCES

1. A. L. Leiner, W. A. Notz, J. L. Smith, A. Weinberger, "An Experimental 16-Bit Memory System Using In-Line Cryotrons," Project Lightning, Ninth Quarterly Progress Report, February 1961, Appendix IV.

2. Proposal for Phase IV of Lightning, "Feasibility Study for a High-Speed Memory," IBM Research Center, February 1961.

APPENDIX X

The Implementation of Logic with Multiple-Control Cryotrons

J. L. Sanborn

ABSTRACT

The use of multiple-control cryotrons permits the implementation of all logic with 2-path circuits. This paper considers several nonexhaustive methods for the synthesis of such circuits. A procedure is also presented for comparing the speeds of the circuits produced by the different methods. A determination is made of the method which appears most promising, and of additional modifications which would enhance the results of that method.

INTRODUCTION

The realization of any large-scale cryotronic information handling system will require a high degree of design automation. Present technologies utilize automation largely for component minimization and bookkeeping. These purposes are also valid for cryotronic systems, but an even more important need for automation is in the placement of cryotrons on the substrate and in the design of individual masks for fabricating the substrates. This is especially important for nonrepetitive sections of the system, such as control.

The size and speed of a cryotronic system are direct functions of the layout. Therefore, an optimum design will be obtained only through interaction between the synthesis and layout portions of the automation program. This precludes the use of semi-exhaustive synthesis methods which are concerned only with a minimum component count.

This paper discusses several synthesis methods, presents a procedure for comparing the speeds of the circuits produced by these methods, and determines the synthesis method which will generally produce the desired results.

MULTIPLE-CONTROL CRYOTRONS

The investigation of a new technology must include comparisons of speed and cost. A simple evaluation of cryotronic circuits shows that competitive speeds probably will not be obtained with the type of cryotron previously used, i. e., with gate width five or six times that of control width. The slow operating speed is a result of both the high inductance introduced by each control line and the small portion of the gate which is driven resistive. For example, consider a circuit in which 80 percent of the total length consists of gates and interconnection lines and the other 20 percent consists of control lines one sixth the gate width. The control lines alone produce 60 percent of the total circuit inductance, and the length of the resistive region is only 1/6 w, where w is the gate width. If the control lines are made equal in width to the gates, the total circuit inductance is halved, and the length of the resistive region (and thus the resistance) is increased six times, for an increase of 12 times over the original circuit speed. The speed may be further improved through the use of in-line cryotrons where the amount of resistance introduced is also proportional to the length of the gate, but where the length may be made much greater than the width. The

increase in speed through the use of in-line cryotrons is particularly striking when the circuits contain long interconnection lines.

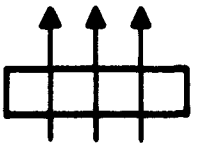
Cryotrons with unity crossings (control width equal to gate width) and in-line cryotrons both have an inherent gain less than one, so that the current-carrying capacity of a gate is less than the control current required to drive all current out of a line containing a similar gate. There exists, however, a range of gate currents from zero to a value less than the current-carrying capacity of the gate where the amount of control current required to introduce resistance increases more slowly than the gate current decreases. This has been termed a differential gain greater than one, and permits the use of bias to obtain operating gains greater than one. The most practical means of applying this bias appears to be through the use of a separate bias line superimposed above the control line. If the amount of bias current required can be made equal to the amount of control current, both lines may be used as control lines so that resistance is introduced only when both signals are present.

A further extension of this technique is the use of three superimposed control lines, where the third line may be either another signal line or a constant-current bias line. If the requirement for introducing resistance is that the algebraic sum of the three control lines be equal to two units of current, no additional restrictions

need be made on the cryotron characteristics beyond those for the simple two-control type. Further lines could also be added provided the line width is very large compared with the total thickness of the gate and control lines. Little additional flexibility is gained, however, unless the discrimination is increased so that two units of current do not introduce resistance but three units of current do. In light of present and future fabrication tolerances, this increased discrimination does not appear justified and is not considered in this paper.

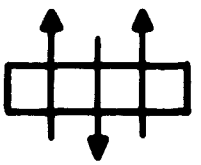
Figure 1 illustrates the types of cryotrons used for the implementations described in this paper. These cryotrons, with at least two input (nonbias) lines, may be either crossed or in-line with no input restrictions except for the final type. In-line cryotrons require that the net control current be unidirectional and opposed to the gate current. Then, in order for the final type to operate properly as an in-line cryotron, the input restriction must be added that if B carries current, either A or C, or both, must also carry current. For example, let $A = X + Y$, $C = X' + W$, and $B = Z$. Since X is always 0 or 1 except for transients, either A or C is always on. The cryotron is resistive for $(X + Y)\cdot(X' + W)\cdot Z' = (XW + X'Y)\cdot Z'$.

Given a set of components such as those in Fig. 1, the next thing to be determined is the type of circuit to be built with the components. In the past, logic with single-control cryotrons has been



x	Resistive	Superconductive
+1	A+B	$A' \cdot B'$
0	$A \cdot B$	$A'+B'$
-2	$A' \cdot B'$	A+B
-3	$A'+B'$	$A \cdot B$
C	AB+AC+BC	$A'B'+A'C'+B'C'$

a. Currents A and B additive.



x	Resistive	Superconductive
+2	$A+B'$	$A' \cdot B$
+1	$A \cdot B'$	$A'+B$
-1	$A' \cdot B$	$A+B'$
-2	$A'+B$	$A \cdot B'$
C	$A \cdot B' \cdot C$	$A'+B+C'$

b. Currents A and B subtractive.

Figure 1 - Multiple-control cryotrons with 2 or 3 distinct inputs.

formed largely through the use of tree circuits, where the supply current is steered to the proper output through the use of a pair of cryotrons similar to relay transfer contacts.[1] The advantage of the tree circuit is the ease with which a function may be hand-implemented. The disadvantage of the tree circuit is the large number of inner loops that result when the number of input variables increases. These loops not only greatly increase the switching time,[2] but also cause large transient currents when the inputs are changed, which in turn may cause false outputs.[3]

An alternate to the tree circuit is a 2-path circuit, similar to a cryotronic flip-flop. The advantages and disadvantages of the 2-path circuit are opposite those of the tree circuit. Switching time is dependent on a single time constant and false outputs cannot occur, since current may flow in only one of two paths, depending on the previous history of the loop and the path in which resistance is introduced. On the other hand, designing with 2-path circuits can be difficult for a person. Since this paper is concerned with machine methods of synthesis, however, this disadvantage is of little consequence.

Single-control cryotrons may be used to produce 2-path circuits with a reset-type logic.[4] This method has a reset cryotron in one side of the circuit and cryotrons in series in the other side for

setting. Reset-type logic may be used well for the design of a set of n-way AND's, such as a selector switch. The difficulty arises in the implementation of OR's, since a timed series of reset pulses is required which will introduce resistance in both sides of the circuit under certain conditions. This problem is alleviated somewhat through the use of multiple-control cryotrons.

Two-path circuits may be easily formed with multiple-control cryotrons. Figure 2 shows one way in which all two-variable functions may be formed in a single level. Note that only one half of the two input loops is required in this particular implementation. This may reduce the area required for interconnections, especially for cryotron circuits constructed in a bifilar manner.

The exclusive-OR function is the only one in two variables which requires four cryotrons in order to be implemented in a single level. This function may also be formed with two cryotrons in two levels provided A'B' and AB (or AB' and A'B) are required in addition to the exclusive OR. Figure 3 illustrates one way of forming this function. In the event that the exclusive-OR function is to be ANDed with a third variable, this may be done with the same number of cryotrons, as shown in Fig. 4. In this case, the pair of inputs available from a previous stage must be AB' and A'B.

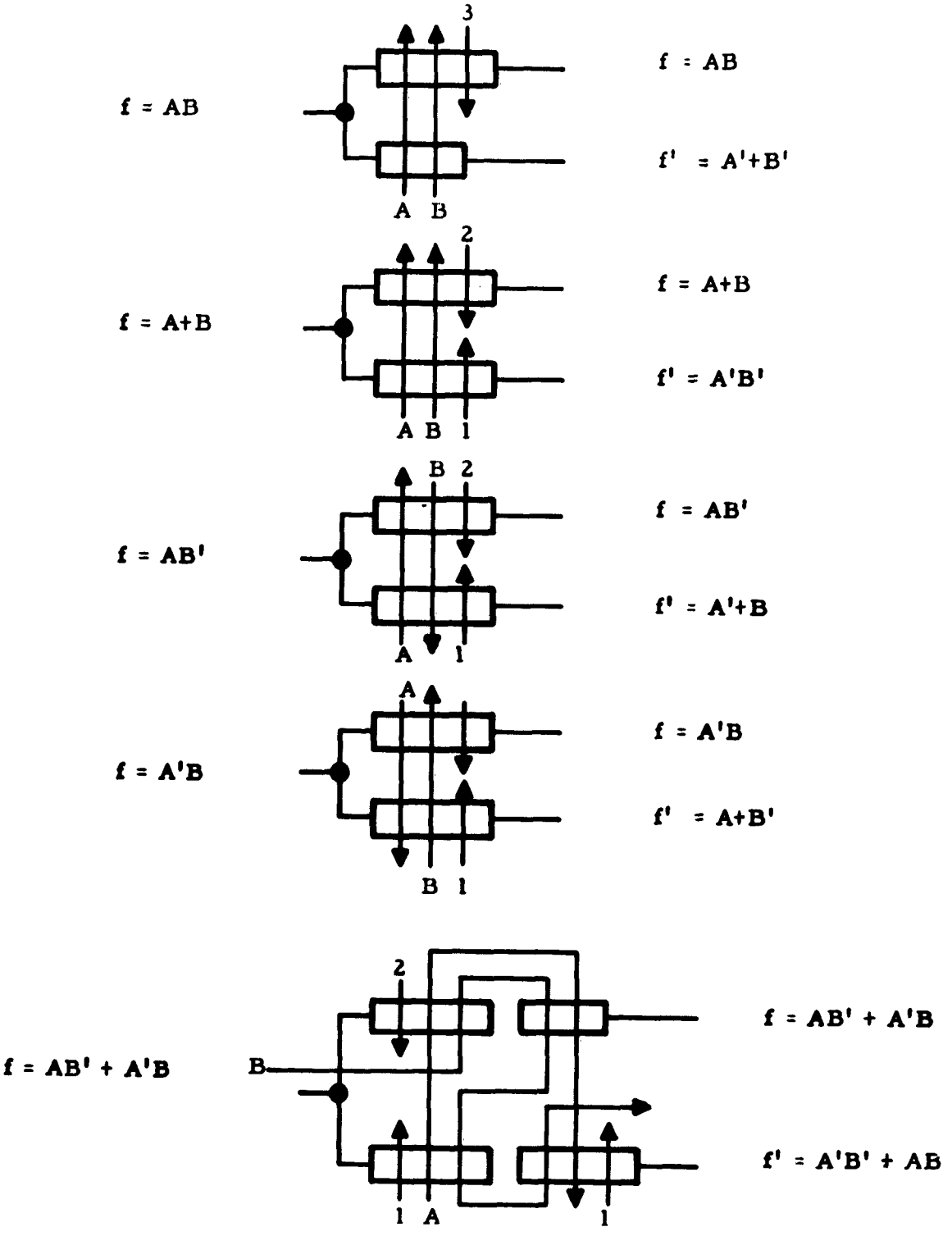


Figure 2 - Complete two-variable functions with multiple-control cryotrons.

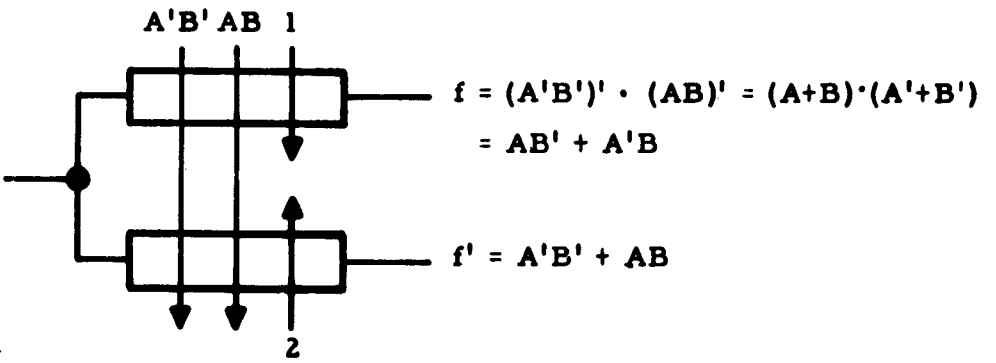


Figure 3 - Alternate implementation of exclusive-OR function.

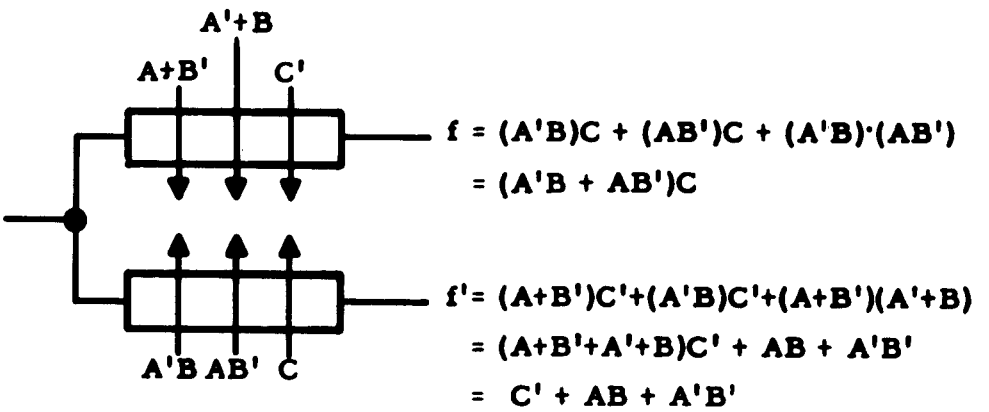


Figure 4 - Implementation of (A'B + AB')C.

This paper will not attempt to prove that 2-path circuits are always better than tree circuits (which they are not), or that they are always faster than tree circuits. Two-path circuits have been adopted as a design philosophy and have been used as a basis for the methods of synthesis to be investigated.

METHODS OF SYNTHESIS

The general problem of synthesizing a Boolean function with a minimum number of components or with a minimum cost has received considerable attention. One such approach has been described by Roth in the IBM Journal of Research and Development.[5] Given a set of building blocks, with a non-negative cost for each block, how can a particular function be realized with a minimum cost? This approach is valid if the cost ratio of component to wire is large and could be modified to produce the fastest circuit if the operating speed is largely dependent on delay in the individual components. Neither of these conditions is satisfied in a cryotronic system.

A cryotronic substrate is fabricated by depositing layers of metal and insulation. The difference between the fabrication of a cryotron gate and an interconnecting wire is in the choice of metal which is deposited at that position. In addition, the effect of a gate

on the operating speed of a circuit is essentially equivalent to an equal length of interconnecting wire. Therefore the optimum implementation of a cryotronic system in terms of cost, size, and speed is dependent on the over-all layout of the system rather than the number of components.

Inasmuch as the selection of a final design must include the physical layout, the method of synthesis should be adapted to the technology in a manner which will simplify the synthesis and still yield the best results for most cases. This simplification is necessary since the actual implementation may require a different synthesis than would be derived simply on the basis of component count.

Direct Implementation

This method represents the most obvious means for implementing a function, and is included largely as a reference point for the other methods.

The function and its inverse are both expressed as minimum sums of products. Each term in these sums is divided in half, then in quarters, and so forth until each subterm contains no more than two input variables. Then each subterm may be implemented by a two-cryotron 2-path circuit. These subterms are then combined in similar circuits until one half of each term in the minimum sums has

been generated. The final function is formed by combining the half-terms as a "two-input AND" cryotron and placing these cryotrons in series.

Direct implementation will generally produce a minimum number of levels for a given function, namely the integer portion of $\log_2(n-1) + 1$, where n is the number of variables. The disadvantage of the method is that the final circuit contains as many cryotrons as there are terms in the minimum sums for the function and its inverse.

Expansion

The standard expansion formula is used in this method to eliminate one variable at a time from the original function. This is performed in the following manner:

$$f(A, B, C, D, \ldots) = A' \cdot f_1 + A \cdot f_2,$$

where f_1 is the result of setting $A = 0$ and $A' = 1$ in the original function and simplifying, and f_2 is the result of setting $A = 1$ and $A' = 0$ in the original function and simplifying. The inverse of the function is simply

$$A' \cdot f_1' + A \cdot f_2'.$$

Then if both f_1 and f_2 are nontrivial functions, i.e., neither 0 nor 1, the final stage consists of a 2-path circuit with two cryotrons in each path.

The reduced functions f_1 and f_2 are next operated on in a similar manner, and the process is repeated until the resultant functions have no more than two variables. Each subcircuit is then a 2-path circuit with no more than four cryotrons, and the function may be formed with no more than n-1 levels, where n is the number of variables.

The expansion method is likely to produce optimum results under any of the following conditions:

1. When the final variable to change is known in advance. If that variable is the first to be eliminated from the function, the proper output from the function will appear after only one stage has switched.

2. When the load on a single variable must be minimised regardless of the effect on the other variables. If that variable is the first to be eliminated, it will control only four cryotrons, regardless of the complexity of the total function.

3. When the subfunctions which result from expanding about a single variable contain less than n-1 variables. This is often the case when only one variable appears in every term of the minimum sum. The additional variables which are eliminated do not have to be the same in the two subfunctions in order for the reduction to be significant.

4. When the subfunctions which result from expanding about a single variable are complements of each other. Although this seems an unlikely condition, five percent of the classes of four-variable functions satisfy the condition.

It should be pointed out that even though the expansion method may not generally produce optimum results, a single application of the expansion formula prior to the general method may prove beneficial when any of the above conditions is met.

Two-Variable Expansion

This is a direct extension of the expansion method in which two variables are eliminated at each level. The means for this elimination is the following formula:

$$f(A, B, C, D, \ldots) = (A'B')\cdot f_1 + (A'B)\cdot f_2 + (AB')\cdot f_3 + (AB)\cdot f_4,$$

where the following substitutions are made in the original formula to determine the f_I's:

	A	A'	B	B'
f_1	0	1	0	1
f_2	0	1	1	0
f_3	1	0	0	1
f_4	1	0	1	0.

Then if all four subfunctions are different and nontrivial, the final stage consists of a 2-path circuit with four cryotrons in each path.

The reduced functions f_1 to f_4 are next operated on in a similar manner, and the process is repeated until the resultant functions contain no more than two variables. Each subcircuit is then a 2-path circuit with two to eight cryotrons in each circuit. The entire function is formed with a maximum number of levels equal to the integer portion of $(n-1)/2 + 1$, where n is the number of variables.

Reduction in the number of cryotrons required for the implementation of a function is obtained when any of the following conditions exists (in approximate order of importance):

1. When two of the subfunctions are equal. This not only reduces the number of subfunctions to be formed, but also reduces the number of cryotrons in that stage and the number of coterms which must be formed. If the coterms corresponding to the subfunctions are unit distant (e. g. AB and AB'), the equivalent logic is performed by a

single cryotron ($A \cdot f_1$). If the coterms are not unit distant (e. g. AB' and A'B), the equivalent logic may usually be performed by a three-control cryotron, where two of the controls are the inverses of the remaining coterms (A +B and A' + B'). In either case, six cryotrons are eliminated in addition to the reduction in the number of subfunctions to be formed.

2. When one of the subfunctions is equal to 1 or 0. This has the effect of reducing the required number of subfunctions and eliminating three cryotrons, one in that stage since the coterm will appear either in the 1 or 0 side but not both, and two in the formation of the coterm, since it may be formed directly in that stage.

3. When one subfunction is the complement of another. This reduces the number of required subfunctions.

4. When two subfunctions consist of single variables and their coterms are unit distant. This is a condition which normally occurs only near the end of a reduction, and eliminates only two cryotrons. For example, (AB')C + (AB)D would require eight cryotrons, but A(B'C + BD) would require only six cryotrons.

Majority Logic

Cohn and Lindaman[6] describe a method in which a Boolean expression is transformed into a majority logic expression.

There are two reasons for believing that this method might prove best for use with cryotronic systems. The multiple-control cryotron is basically a threshold device having a threshold of two. Therefore if the cryotron has three controls, the majority function may be formed with a single cryotron. In addition to requiring only a single cryotron, the inverse of a majority function could be formed by using the inverse of the inputs to the original function.

This method performs the decomposition through the use of one of the following formulas:

$$\begin{aligned} f(A, B, C, D, \ldots) &= (A\#B\#f_{ab'})\#(A'\#B'\#f_{ab'})\#f_{ab} \\ &= (A\#B'\#f_{ab})\#(A'\#B\#f_{ab})\#f_{ab'} \\ &= (A\#B\#f_{ab'})\#(A'\#B\#f_{ab})\#B' \\ &= (A'\#B\#f_{ab})\#(B'\#C\#f_{bc})\#(C'\#A\#f_{ca}), \end{aligned}$$

where $A\#B\#C = A \cdot B + A \cdot C + B \cdot C$, the standard majority function, and where f_{ab} is formed by setting $B = A$ and $B' = A'$, and $f_{ab'}$ is formed by setting $B = A'$ and $B' = A$, etc. The final stage is then a 2-path circuit containing a single three-control cryotron in each path with inputs as indicated in the formulas.

The subfunctions are next operated on in a similar manner, and the process is repeated until the resultant functions have no more than two variables. Each subcircuit is then a 2-path circuit with two

cryotrons (with the exception of the exclusive-OR function of two variables, if required), and the function may be implemented with no more than 2(n-1) levels, where n is the number of variables.

The choice of which of the four formulas to use for a given function is largely dependent on the complexity of the subfunctions. The complexity can best be measured by the number of logic levels required for implementation. If the two subfunctions differ in complexity, one of the first two formulas should be used, so that the more complex subfunction serves as the third input to the final cryotron. If the two subfunctions are essentially equal, any of the first three formulas will produce comparable results. The fourth formula may produce better results when the number of variables increases, but no four-variable function has been found for which this is true.

EVALUATION METHODS

Methods for implementing a given function with cryotrons may be compared in two ways - in size and speed. In order to make a rigorous comparison, a complete layout is required. An acceptable preliminary comparison may be made by considering the number of cryotrons required and the character of the circuit.

The size of a cryotronic system is largely a function of the substrate area required, since the height of a cryotron structure is small compared to its other dimensions. The substrate area is in turn a function of the number of cryotrons and the interconnections required between functional blocks. Since this paper is concerned only with implementing a single function at a time, the number of cryotrons represents a first approximation to the size of the block.

The absolute speed of a cryotronic system is a function of parameters which are determined during the fabrication and operation of a specific system as well as the specific layout. When comparing alternate methods of performing the same function, the comparison need only be in relative rather than absolute terms.

The switching speed of a circuit is of the form nL/R, where n is determined by the percentage of current which must be switched before resistance is introduced in the next circuit and R is the amount of resistance introduced in the present circuit. For a 2-path loop, the effective inductance L is simply the total inductance of the circuit. Assuming a constant n and R for the entire block, which amounts to assuming the same operating characteristics for all cryotrons, L is a measure of the speed of 2-path circuits.

The methods of synthesis described all produce a set of 2-path circuits. When an input variable changes, a chain reaction is

started which results in the output of the block either remaining the same or changing. Assuming that the individual inputs are equally loaded outside the block, the maximum switching speed is determined by the maximum inductance which must be affected. With 2-path circuits, this inductance consists of the inductive load placed on the input by the block and the sum of the inductances of the individual circuits comprising the chain.

The inductance of a circuit is composed of two main parts - that introduced by the gates and controls, and that introduced by lines connecting the gates and controls. The interconnection lines contribute a nonnegligible amount of inductances to the circuit. It appears reasonable to assume, however, that the length of interconnection line required is proportional to the number of gates and controls in the circuit. This permits circuits to be compared relatively on the basis of the inductance introduced by their gates and controls.

The inductance of a superconducting line has been expressed precisely as a function of the basic parameters of the materials concerned.[7] These formulas may be simplified considerably if the thickness of the line is greater than or equal to four times the effective penetration depth of the line material. If, in addition, the cryotrons are considered to be dimensionally equal and insulation thicknesses are equal, the inductance is a function only of the thickness and

materials in and between the line and the ground plane.

A cross-section view of an in-line cryotron is shown in Fig. 5. Based on the preceding assumptions, the inductance of a line may be expressed as a constant times another term which will be called the electrical separation between the line and the ground plane. Table I lists the formulas to be used, depending on the specific line and the state of the cryotron gate. Note that the electrical separation may be greater or less than the physical separation.

The electrical separations for the various lines are shown in Table II, based on a presently attainable set of parameters, expressed both in angstroms and normalized with respect to the electrical separation of the gate. Either set of values could be used for computer evaluation, but both become unwieldy for hand calculations. For this reason, the integer values of 1, 2, 3, and 4 have been chosen to represent the respective lines of the cryotron. The percentage deviation of these integers from the actual values is also shown in Table II.

The chain length or path length may now be represented as the sum of the normalized electrical separations of the individual gates and controls in the chain. The speed of a block is the maximum path length from input to output, and the best implementation in terms

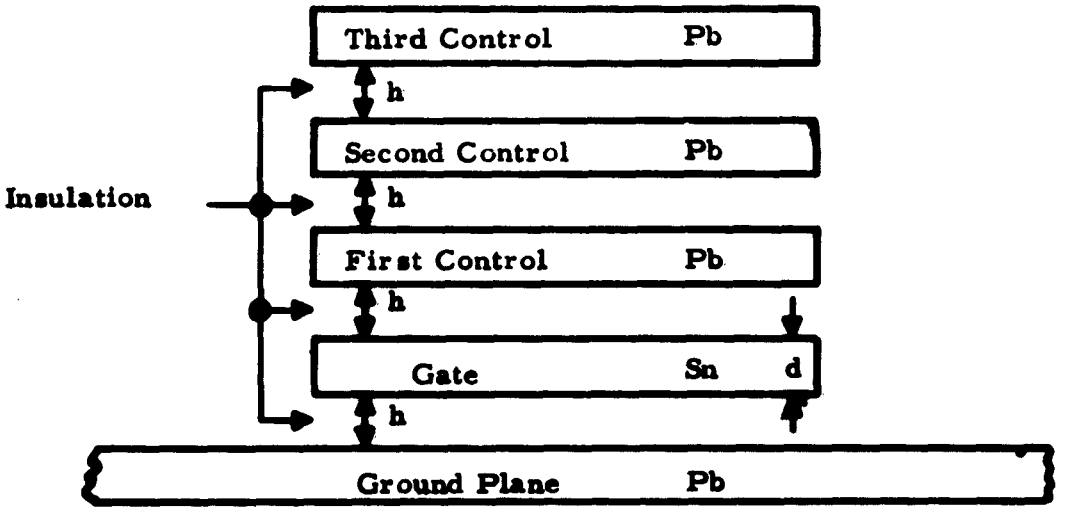


Figure 5 - Cross-section of in-line cryotron.

Table I - Formulas for Electrical Separation between Line and Ground Plane[a]

	Gate Superconductive	Gate Resistive
Gate	$\lambda_{sn} + \lambda_{pb} + h$	$1/3d + \lambda_{pb} + h$
First Control	$2\lambda_{sn} + 2\lambda_{pb} + 2h$	$d + 2\lambda_{pb} + 2h$
Second Control	$2\lambda_{sn} + 4\lambda_{pb} + 3h$	$d + 4\lambda_{pb} + 3h$
Third Control	$2\lambda_{sn} + 6\lambda_{pb} + 4h$	$d + 6\lambda_{pb} + 4h$

[a] λ = effective penetration depth
h = insulation thickness
d = gate thickness.

Table II - Electrical Separation between Line and Ground Plane[a]

	Gate - Superconductive			Gate - Resistive		
	Actual	Normalized	Percent Deviation	Actual	Normalized	Percent Deviation
Gate	7000	1.00	0	7500	1.00	0
First Control	14 000	2.00	0	17 000	2.26	-13
Second Control	20 000	2.86	+5	23 000	3.06	- 2
Third Control	26 000	3.71	+7	29 000	3.86	+ 4

[a] In angstroms. Assumed values are $\lambda_{sn} = 1500$
$\lambda_{pb} = 500$
$h = 5000$
$d = 6000$.

of speed is determined by the minimum of the maximum path lengths of the various blocks.

SAMPLE PROBLEM

The best way to understand the various methods of synthesis is through a sample problem. All of the methods require operation on the given function with several combinations of the input variables. If the function is partially symmetric, i. e., if the function is not changed when one variable is interchanged either with another variable or with the inverse of another variable, the number of combinations which must be tried is reduced.

The following four-variable function was chosen to illustrate the methods of synthesis:

$$f = A'B'D' + A'C'D' + B'C'D' + BCD,$$

$$f' = BC'D + B'CD + BCD' + AD'.$$

This function is symmetric about B and C, as can be seen by inspection.

Direct Implementation

The direct implementation of a four-variable function would normally require three divisions into two sets of two variables.

Since B and C may be interchanged, two of the divisions are equivalent and only one of them need be performed:

AB/CD $f = (A'B')D' + A'(C'D') + B'(C'D) + B(CD)$,

$f' = B(C'D) + B'(CD) + B(CD') + AD'$.

AC/BD equivalent to AB/CD.

AD/BC $f = (A'D')B' + (A'D')C' + (B'C')D + (BC)D$,

$f' = (BC')D + (B'C)D + (BC)D' + AD'$.

Since this method is included mainly for reference, the choice of the division to be implemented was made on the basis of the number of cryotrons. This may be readily determined for four variables by observing that the final stage of a particular function will have the same number of cryotrons regardless of the division. The minimum number of cryotrons is therefore determined by the minimum number of different subfunctions within the parentheses. There are five different subfunctions in each division so that the same number of cryotrons is required. Note that in f' for the division AD/BC, the term AD' is not in parentheses since it may be formed with a single cryotron in the second stage versus two cryotrons in the first stage.

A block diagram of the implementation of division AD/BC is shown in Fig. 6a. The block diagram has been drawn in terms of

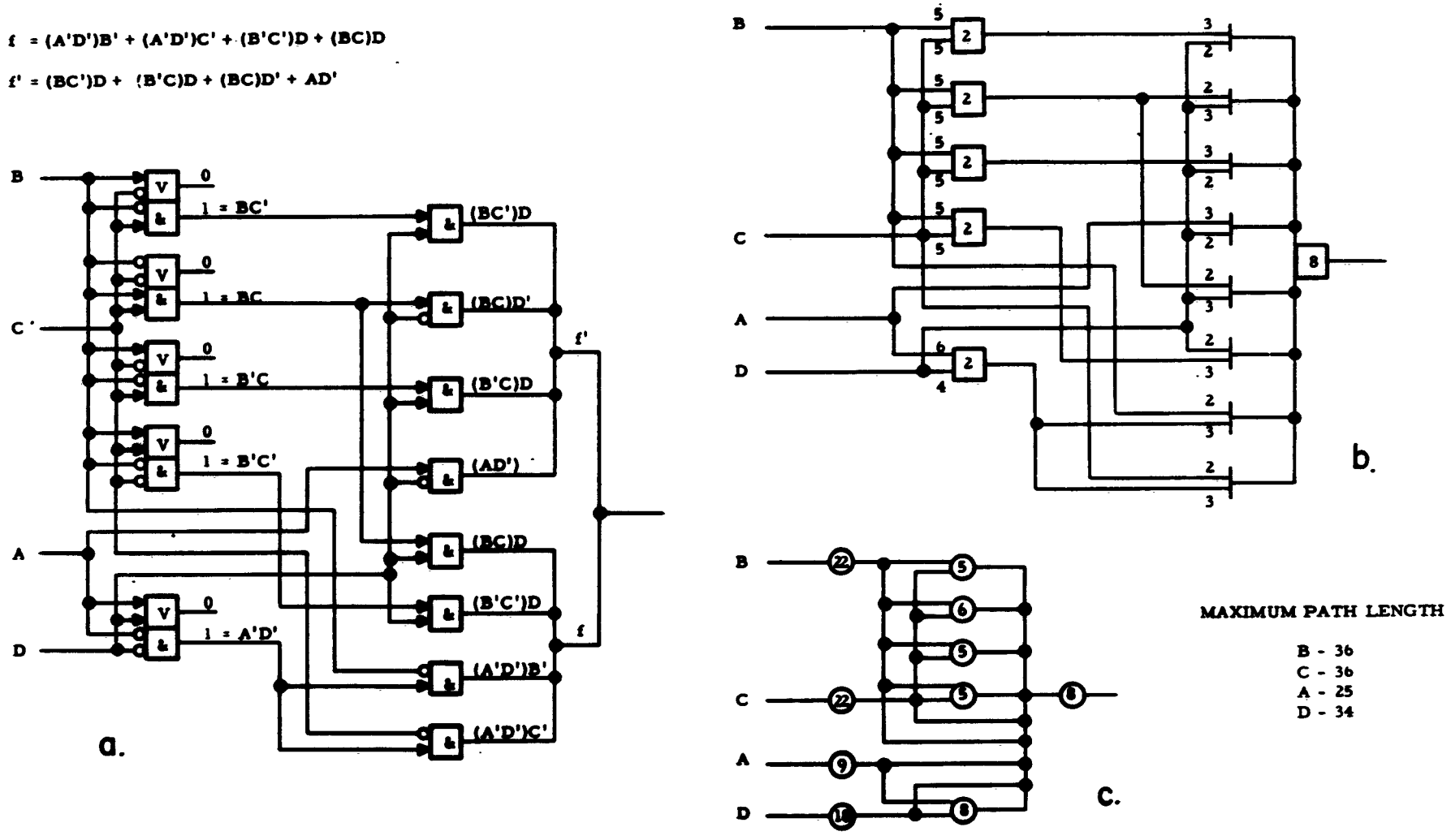


Figure 6 - Direct implementation.

positive logic in order to clarify the logic flow, whereas the use of cryotrons is a type of negative logic. Lines with arrows leading into a block are direct inputs and lines with circles indicate that the inputs are used to overcome bias lines.

Figure 6b is a compressed block diagram which is concerned not with the actual logic being performed but with the origin and destination of the cryotron inputs. A two-input AND, for example, is represented by two lines leading into a single block containing the number 2. This number represents both the number of cryotrons in the circuit and the sum of the electrical separations of the individual cryotron gates. Since the circuit contains two cryotrons with two inputs each, there are two first controls and two second controls. The electrical separations associated with these controls must total 10 and must be distributed between the two inputs either 5/5 or 6/4. If each cryotron had three inputs, the total separation would be 18 and the possible distributions among the three inputs would be 6/6/6, 5/6/7, 5/5/8, and 4/6/8. The determination of whether a control should be a first or second control is largely a matter of trial and error. The desired result is to reduce the maximum path length as much as possible. The figures in this paper represent what is felt to be the best result for each implementation. The number of cryotrons used in an implementation may be easily

determined by summing the squared numbers in this diagram. Eighteen cryotrons are used in this example.

The maximum path length is determined with the aid of a further compression shown in Fig. 6c. The circled number associated with each input is the sum of the separations of the controls supplied by that input. Each inner circled number is the sum of the separations of the gates and controls for one circuit. The circled number to the right of the diagram is the sum of the separations of the gates in the final stage. The maximum path length for each input is found by adding the circled numbers for all possible paths between that input and the output. The maximum number represents the maximum path length. The largest path length thus determined represents the speed of the implementation, in this case, 36.

Expansion

The application of this method to a four-variable function would normally require four initial expansions. Since B and C may be interchanged, two of the expansions are equivalent and a total of three expansions is sufficient. The second level expansions shown below were chosen so as to minimize the number of cryotrons:

A $f = A'(B'D' + C'D' + B'C' + BCD) + A(B'C'D + BCD)$

$= A'[D'(B' + C') + D(B'C' + BC)] + A[D(B'C' + BC)]$.

B $f = B'(A'D' + C'D) + B(A'C'D' + CD)$

$= B'[D'(A') + D(C')] + B[D'(A'C') + D(C)]$.

C equivalent to B.

D $f = D'(A'B' + A'C') + D(B'C' + BC)$

$= D'[A'(B' + C')] + D(B'C' + BC)$.

All three of the expansions should be evaluated in order to be completely rigorous. Experience with the method has shown, however, that the following rules of thumb hold for four variables:

1. If two minimum implementations differ by more than two cryotrons, the lower of the two will always produce a faster circuit.

2. If two minimum implementations differ in the number of logic levels, two should be subtracted from the cryotron count of the implementation with fewer levels before Rule 1 is applied.

Since all three expansions require three logic levels, which may be determined from the level of brackets, only Rule 1 applies. Unfortunately, determining the number of cryotrons is not a straightforward problem. This is due largely to the fact that the inverse of any function is available and that three-control cryotrons will often reduce the total cryotron count. The minimum cryotron counts for

the three expansions are A - 14, B - 14, and D - 10. Since D requires four less cryotrons than A or B, only D need be evaluated. This is shown in Fig. 7 and results in a maximum path length of 30. In Fig. 7b, two lines are used in the final stage to indicate a pair of cryotrons with the same inputs.

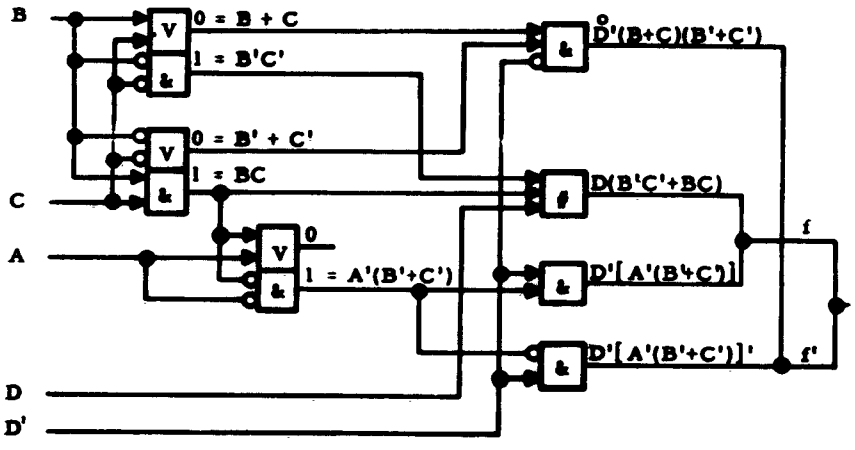
The first expansion in D resulted in two subfunctions. The second subfunction contains only two variables, which can always be formed in one level. The first subfunction contains three variables, but consists of two terms with no more than two variables each. This can also always be formed in one level, but in this case the number of cryotrons for the over-all function is increased. This implementation is shown in Fig. 8 and has a maximum path length of 27. Thus the addition of two cryotrons has reduced the over-all path length by 10 percent.

Two-Variable Expansion

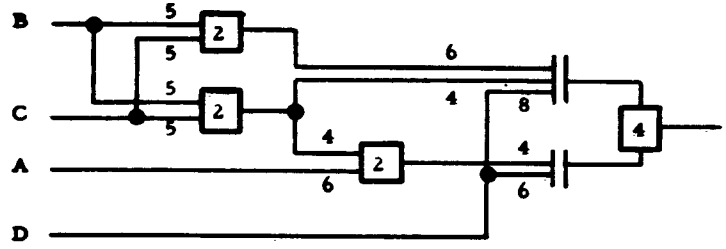
This method requires the evaluation of four subfunctions for each combination, and thus the detection of symmetries is even more important. In the example under consideration, the symmetry of B and C reduces the combinations to four and the subfunctions to be evaluated to sixteen.

f = D'[A'(B'+C')] + D(B'C'+BC)

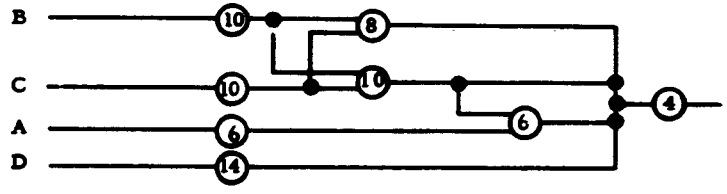
f' = D'[A'(B'+C')]' + D⊙(B+C)(B'+C')



a.



b.



c.

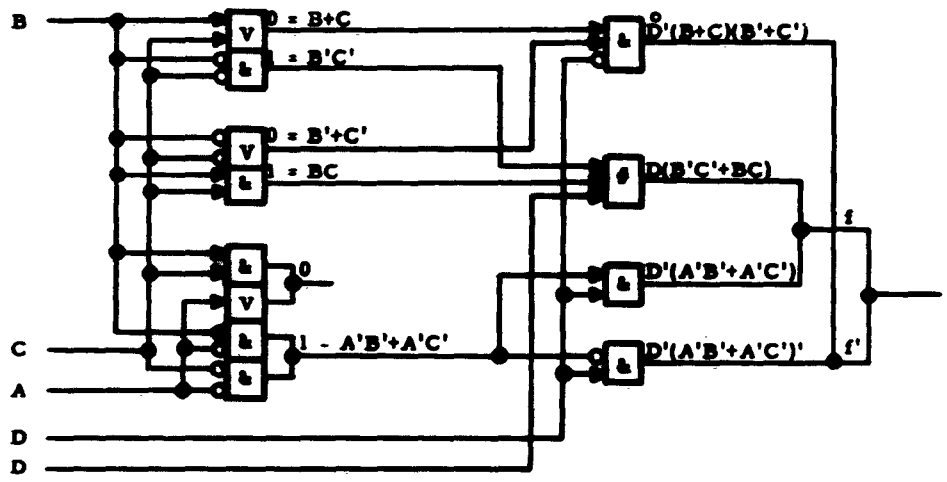
MAXIMUM PATH LENGTH

B - 30
C - 30
A - 16
D - 18

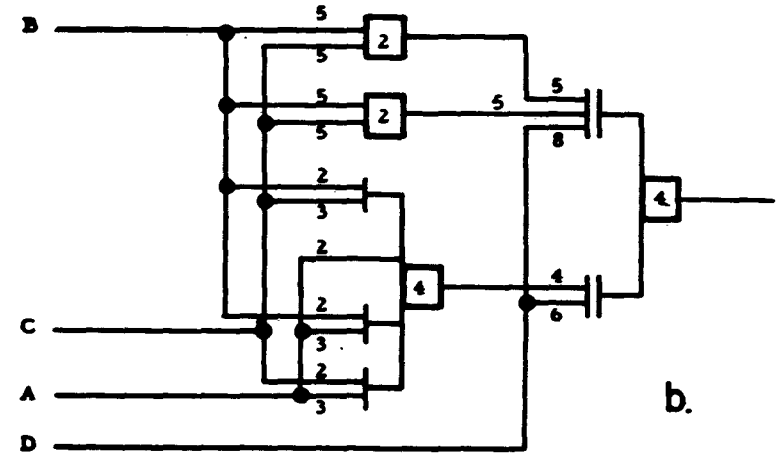
Figure 7 - Expansion - Case 1.

f = D'(A'B'+A'C') + D(B'C'+BC)

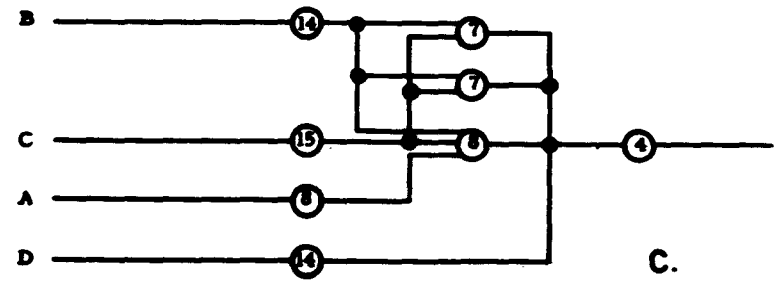
f' = D'(A'B'+A'C')' + D'(B+C)(B'+C')



a.



b.



c.

MAXIMUM PATH LENGTH

B - 26
C - 27
A - 20
D - 18

Figure 8 - Expansion - Case 2.

AB $f = (A'B')(D' + C') + (A'B)(C'D' + CD) + (AB')(C'D) + (AB)(CD).$

AC equivalent to AB.

AD $f = (A'D')(B' + C') + (A'D)(B'C' + BC) + (AD')(0) + (AD)(B'C' + BC)$

$= (A'D')(B' + C') + D(B'C' + BC) + (AD')(0).$

BC $f = (B'C')(A' + D) + (B'C)(A'D') + (BC')(A'D') + (BC)D$

$= (B'C')(A' + D) + (B'C + BC')(A'D') + (BC)D.$

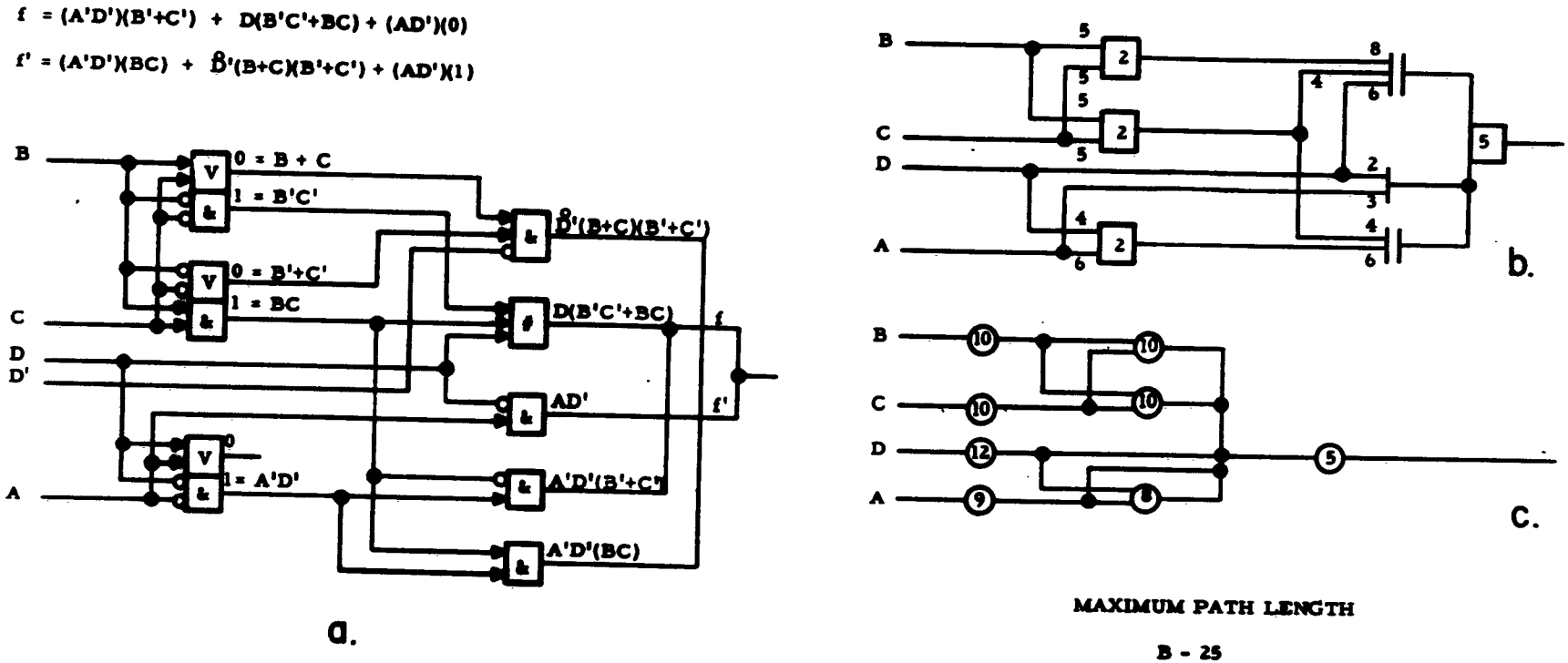
BD $f = (B'D')A' + (B'D)C' + (BD')(A'C') + (BD)C$

$= B'(A'D' + C'D) + (BD')(A'C') + (BD)C$

$= (B'D')A' + (BD')(A'C') + D(B'C' + BC).$

CD equivalent to BD.

The rules of thumb described for the expansion method also apply for the two-variable expansion method. Similarly, the determination of the minimum number of cryotrons required is not a trivial problem. The minimum cryotron count for each of the four expansions is AB - 16, AD - 11, BC - 12, and BD - 14. Since AB and BD exceed the minimum implementation AD by more than two cryotrons, they need not be evaluated. The implementation of AD is shown in Fig. 9 and results in a maximum path length of 25. The implementation of BC is also required but not shown, and results in a speed of 32.



MAXIMUM PATH LENGTH

B - 25
C - 25
D - 25
A - 22

Figure 9 - Two-variable expansion.

Majority Logic

There are six combinations of four things taken two at a time. In this particular example, however, the symmetry of B and C reduces the unique combinations to four:

AB $f_{ab} = A'D' + A'C' + ACD,$

$f_{ab'} = A'C'D' + AC'D + A'CD.$

AC equivalent to AB.

AD $f_{ad} = A'B' + A'C' + B'C' + ABC,$

$f_{ad'} = A'B'C' + A'BC.$

BC $f_{bc} = A'B' + D,$

$f_{bc'} = A'D'.$

BD $f_{bd} = A'B' + BC,$

$f_{bd'} = A'C' + B'C'.$

CD equivalent to BD.

Determining the number of cryotrons for each combination is even more complicated than in the expansion method since it is now a multiple-output problem. The minimum cryotron counts are AB - 12, AD - 10, BC - 6, and BD - 8. Then the two possible best implementations are BC and BD. Since f_{bc} and $f_{bc'}$ may each be obtained in a

single level, there are three possible implementations with three levels and twelve cryotrons:

$$f_1 = [B\#C\#(A'D')]\#[B'\#C'\#(A'D')]\#(A'B' + D),$$

$$f_2 = [B\#C'\#(A'B' + D)]\#[B'\#C\#(A'B' + D)]\#(A'D'),$$

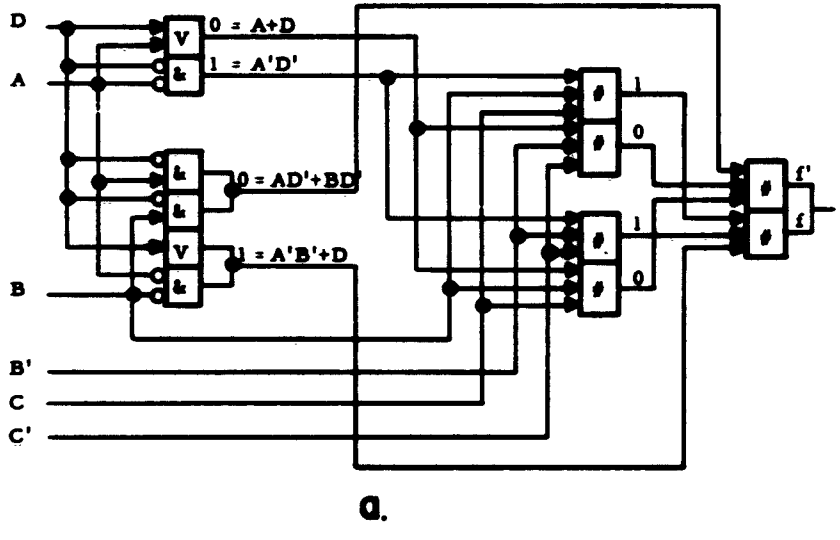
$$f_3 = [B\#C\#(A'D')]\#[B'\#C\#(A'B' + D)]\#C'.$$

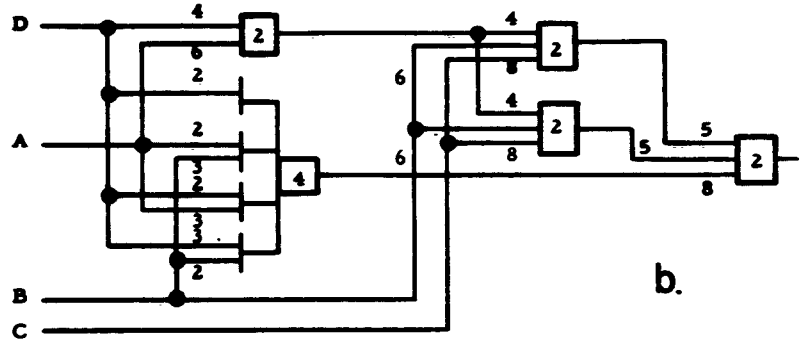
The implementation of f_1, shown in Fig. 10, results in a maximum path length of 31; f_2, 38; and f_3, 31. With the addition of four cryotrons, f_3 may be implemented with a maximum path length of 30. This reduction is not felt to be sufficient to justify the additional cryotrons.

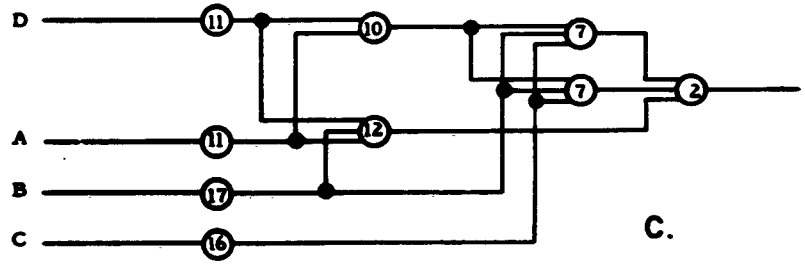
BD may be implemented with only two more cryotrons than BC, and should therefore be evaluated. The best speed which could be obtained for BD was 38, approximately 25 percent slower than BC.

SUMMARY

The only previous synthesis method developed for cryotrons is the minimization of tree circuits.[8] This method was programmed for the IBM 704 and the minimum circuits found for all 221 classes of functions of four variables. These classes represent all functions of four variables in that a function not specifically included is either the







MAXIMUM PATH LENGTH

D - 30
A - 30
B - 31
C - 25

Figure 10 - Majority logic.

inverse of a class or is obtained by altering the input variables by permutation, negation, or both.

Because none of the methods described here has been programmed, only 12 of the classes of four-variable functions were considered, approximately 5 percent of the total. The results are summarized in Table III. The figures under the heading "Tree" are the minimum number of cryotrons as determined by the 704 program and the number of permutations of the input variables (out of 24) for which this minimum is obtained. Listed in the next column are the minimum cryotron counts for direct implementation and the speed (in terms of maximum path length) for that implementation. The possibility of reducing the maximum path length by the addition of more cryotrons was investigated for the remaining three methods. Thus for some combinations of class and method, two sets of figures are shown. One set of figures (minimum cryotron count and speed) is the result of implementing the function with that method with the fewest cryotrons, regardless of the speed of the circuit. The second set of figures (maximum cryotron count and speed) is the result of determining the fastest implementation with that method, regardless of the number of cryotrons.

The method which produces the fastest circuit and the circuit with the fewest cryotrons for each example is also shown in

Table III - Summary of Twelve Four-Variable Functions

	Tree		Direct		Expansion				Two-Variable Expansion				Majority Logic			
Class	Cry.	Perm.	Cry.	Speed	Cry.		Speed		Cry.		Speed		Cry.		Speed	
					Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
11	14	8	18	26	13	14	33	30	(11)	13	27	△25	12	--	--	29
27	16	24	22	39	14	--	--	33	(12)	16	33	△29	(12)	14	36	35
44	14	8	21	38	(12)	14	36	31	(12)	14	32	△30	(12)	14	41	34
62	12	12	11	△24	12	--	--	26	12	--	--	27	(10)	12	28	26
82	10	1	13	27	(10)	--	--	△21	(10)	--	--	△21	14	--	--	36
107	16	24	30	49	(16)	20	38	37	(16)	--	--	△33	18	--	--	41
133	12	2	18	36	(10)	12	30	27	11	--	--	△25	12	14	31	30
156	12	1	22	40	(12)	14	32	30	(12)	14	33	△28	14	18	41	35
158	14	8	21	40	14	18	35	34	(12)	--	--	△29	(12)	14	31	30
178	12	5	14	31	(10)	12	30	29	(10)	12	31	△26	14	--	--	35
196	14	4	25	42	16	--	--	37	(14)	--	--	△30	(14)	--	--	35
219	10	2	16	34	(8)	--	--	△24	12	--	--	27	14	20	41	35

△ - Maximum speed regardless of cryotron count.

○ - Minimum cryotron count regardless of speed.

Table III, with duplicate entries in case of ties. The two-variable expansion method produces the fastest circuit in all but two cases. Class 62 has an extremely simple minimum sum, so that only two subterms need be formed in the direct implementation method. Class 219 is one of the special cases already discussed in which the two subfunctions developed by the expansion method are complementary. This will normally require fewer cryotrons, and in this case produces the fastest circuit. The method which implements the functions with the minimum cryotron count is not as clear-cut, although the two-variable expansion came out slightly ahead of the expansion method in the examples considered.

The poor showing of the majority logic caused a more thorough investigation of this particular case. Specifically, functions of the following forms were considered:

$$f = f_1 \# f_2 \# f_3,$$

$$f = A' \cdot f_1 + A \cdot f_2,$$

where any or all of the subfunctions could be majority functions of the input variables. It was felt that circuits of this type, which require only two levels, should produce the fastest circuits. Twenty-four of the classes generated in this manner were faster than could be developed by using the other methods. Of the 24 classes, 6 could be

derived from the majority logic formulas described, 5 were of the second form and could be derived by a combination of the expansion and majority methods, and 13 could not be derived at all. This indicates that either a table lookup procedure should be used to detect these classes, or that the majority logic method should be modified to include them.

CONCLUSIONS

The two-variable expansion method will produce the fastest circuit 75 to 85 percent of the time. This method is almost as effective in producing circuits with the fewest cryotrons, closely followed by the expansion method. Partial symmetry between input variables should be detected, since its presence reduces the number of operations to be performed. The detection of the special cases best handled by either the expansion method, the majority logic method, or a combination of both, would produce a synthesis program which could always develop the best implementation.

ACKNOWLEDGMENT

The author has been helped in the preparation of this material by discussions with many persons, and in particular Dr. W. C. Carter, Dr. J. H. Griesmer, Dr. N. H. Meyers, N. Rochester, and E. H. Sussenguth.

REFERENCES

1. D. A. Buck, "The Cryotron - A Superconductive Computer Component," Proc. Inst. Radio Engrs., April 1956.

2. M. A. Epstein, "Transients for a Cryotron Tree Switch," MIT Lincoln Laboratory Report 6M-4226, March 1956.

3. W. C. Carter and J. L. Sanborn, "Simulation of the Subtractor-Ring Circuit," Appendix VI, this report.

4. N. Rochester, Unpublished work, June 1960.

5. J. P. Roth, "Minimization over Boolean Trees," IBM Journal of Research and Development, November 1960.

6. M. Cohn and R. Lindaman, "Axiomatic Majority-Decision Logic," Inst. Radio Engrs. Trans. on EC, March 1961.

7. N. H. Meyers, "Inductance in In-Line Superconducting Structures," IBM Memorandum, EMRc 440, September 1960.

8. E. H. Sussenguth, "An Algorithm for Automatic Synthesis," Project Lightning, Fifth Quarterly Progress Report, April 1960.

APPENDIX XI

Guides for the Organization of Machines Using Multiplexed Equipment

W. C. Carter

ABSTRACT

Circuit operation at kilomegacycle rates causes the signal propagation delay between units to become important. To reduce the importance of this delay, the system may be divided into relatively self-sufficient subunits operating concurrently and in collaboration, as far as possible, on the same problem.

Quantitative estimates of the efficiency of such multiplexed units operating concurrently will be provided from analysis of a mathematical model. These estimates for particular models may be used to check a designer's intuition. Secondly, these computed efficiency estimates will be used to propose a set of intuitive guides for the design of complex multiplexed equipment.

INTRODUCTION

The ultra-high-speed machines of the future will presumably be composed of very large numbers of switching units operating at kilomegacycle rates. Operation at these rates causes signal propagation delay between the units to become important. It is thus desirable to package the units as compactly as possible, and have little control and data flow between units.

To reduce the necessary signal propagation the system may be divided into relatively self-sufficient subunits. This leads to the concept of a network of concurrently operating computers all tied together and working in collaboration - as far as possible - on the same problem. Intuitively it would appear that this multiplexing would lead to greater speed of problem solution, but closer analysis reveals pitfalls.

The efficiency of multiplexed units of a computer operating concurrently on a problem will be analyzed using a mathematical model to provide quantitative estimates. These computed quantitative estimates for particular models may be used to check a designer's intuition. As a second step in analysis, these computed efficiency estimates will be used to propose a set of intuitive guides for the design of complex multiplexed equipment.

DESCRIPTION OF THE BASIC MATHEMATICAL MODELS

A mathematical model of a physical process may be guided by the following criteria. First, its action must describe the process as accurately as possible. Second, the model can be manipulated to derive quantitative results not obvious from the description. The more complicated the process, the more difficult the definition of the model. In defining the following models, the derivation of quantitative results has been stressed at the expense of accurate description.

These models of multiplexed computer systems contain two main parts. The first part consists of sets of distinct types of independently operating processors, each processor type being controlled by a distinct type of command. The number of types and the number of units of each type vary with the model. The second part consists of sets of strings of the commands and mechanisms for scanning and delivering them to the appropriate unit. Only the probability distribution of the commands in the set of strings is assumed known.

To handle this notationally, each model contains m types of processing units, and a particular type is designated by i, where i varies from 1 to m. If a_i, $i = 1, \ldots, m$, is a positive integer or zero, then a_i will denote the number of processing units of type i in a model. Each particular model is thus defined by an m-tuple $a = (a_1, a_2, \ldots a_m)$.

The range of indexes specifies the types of processing units in the model, and the magnitude of the components specifies the number of processing units of the type defined by the index. Each particular unit of type i is controlled by a command of type I_i where i varies from 1 to m. A particular unit of type i, the unit i_0, upon receiving a command I_{i0}, initiates an action which lasts one time step.

The commands I_i occur in strings, and the probability distribution $\Phi(I_1, \ldots I_m)$ is known.

SEQUENTIAL SCAN MODEL

In the sequential-scan model, the commands I_i are scanned sequentially. At each step it is determined if a processor of type i is available. If a processor is available, the scan continues; otherwise it stops. After the scan stops, all processing units which have received commands process them. The processing and scanning are overlapped. For any set of processing units, the number of operating units is determined by the contents of the scanned string.*

For example, the model $(2_1, 1_2)$ contains two processors of type 1 and one processor of type 2. The command string

$$I_1 I_1 I_2 I_2 I_1 I_2 I_1 I_1 I_2 I_1 I_1 I_1 I_1 I_1 I_1 \ldots$$

* This is very similar to a model originally proposed by M.C. Andrews and W. L. Duda.

will be scanned and processed in the groups shown

$$I_1I_1I_2 \quad I_2I_1 \quad I_2I_1I_1 \quad I_2I_1I_1 \quad I_1I_1 \quad I_1I_1 \ldots$$

It is clear that varying numbers of the processing units are used. The results of the previous scan may be written as the 2-tuples

$$(2, 1), \ (1, 1), \ (2, 1), \ (2, 1), \ (2, 0), \ (2, 0)$$

where the first position denotes the number of steps of type 1 processed, and the second, the steps of type 2.

The efficiency E' for a particular scan of a particular model is defined as

$$E' = \frac{\text{the number of program steps processed}}{\text{the maximum number which could be processed}}. \quad (1)$$

In the example above, the consecutive values of E' are as follows:

Time step	1	2	3	4	5	6 ....
E'	1	2/3	1	1	2/3	2/3 ....

In this example, the set of all possible 2-tuples produced by the scan, the corresponding scan sequences, and the values of E' are listed below.

	Scan Sequence		
2-tuple	Instruction Scanned	Next Instruction	E'
(2, 0)	$I_1 I_1$	I_1	2/3
(2, 1)	$I_2 I_1 I_1$	any	1
	$I_1 I_2 I_1$	any	
	$I_1 I_1 I_2$	any	
(1, 1)	$I_1 I_2$	I_2	2/3
	$I_2 I_1$	I_2	
(0, 1)	I_2	I_2	1/3

If the probabilities of occurrence of the possible scan sequences are written Pr((2, 0)), Pr((2, 1)), etc., then the expected efficiency of the model for this example is

$$E = 2/3 \Pr((2,0)) + \Pr((2,1)) + 2/3 \Pr((1,1)) + 1/3 \Pr((0,1)). \quad (2)$$

To generalize from the example, let the results of each sequential scan be expressed as an m-tuple $a = (a_1, a_2, \ldots, a_m)$, where a_j represents the number of processing units of type j which will be used in the next time step. Let it be assumed that the probabilities of occurrence of the possible scan sequences, Pr(a), can be derived from the probability distribution Φ of the various instruction types, and from assumptions R concerning the conditional probabilities of occurrence of instructions in the scan. By letting A represent the

set of all possible m-tuples $a = (a_1, a_2, \ldots, a_m)$ which can result from a sequential scan, and using $E'_a(\alpha)$ for the efficiency of the model α for a particular scan a, equation (2) can be written for the general case as

$$E(\alpha, \Phi, R) = \sum_{a \in A} E'_a(\alpha) Pr(a). \tag{2a}$$

If the norm of a vector a is defined as the sum of its components and written as $||a||$, equation (1) can be rewritten as

$$E'_a(\alpha) = \frac{||a||}{||\alpha||},$$

so that the expected efficiency of the model becomes

$$E(\alpha, \Phi, R) = \frac{1}{||\alpha||} \sum_{a \in A} ||a|| \, Pr(a). \tag{3}$$

Define the maximum value of E as

$$EMAX(\alpha, \Phi, R) = \max_{\phi \text{ in } \Phi} E(\alpha, \Phi, R) \tag{4}$$

and the average efficiency as

$$F(\alpha, \Phi, R) = \int_{-\infty}^{\infty} E(\alpha, \Phi, R) \, d\Phi. \tag{5}$$

SKIPPING-SCAN MODEL

In the skipping-scan model, the commands I_i are held before processing in a store containing n places. The commands are scanned beginning with the first place. At each step it is determined if a processor of type i is available. If a processor is available, the command is marked for processing. Commands which cannot be processed are retained in the store. The scan continues over all n commands.

All marked commands are processed during the next step. All others are moved to the lowest numbered place possible. In addition, the store is refilled sequentially from the program string, and then the store is rescanned. Processing of commands and the refilling of the n-store are simultaneous. Each model may be defined by the (m+1)-tuple $(a|n) = (a_1, a_2, \ldots a_m|n)$ where the a_i's are as defined previously and n is the number of storage cells. It is assumed that $n \geq ||a||$.

For example, the model $(2_1, 1_2|4)$ contains two processors of type 1, one of type 2, and a store of four cells. The command string

$$I_1 I_1 I_2 I_2 I_1 I_2 I_1 I_1 I_2 I_1 I_1 I_1 I_1 I_1 I_1 \ldots\ldots$$

of the previous example is processed as follows:

contents of store	processed commands
$I_1I_1I_2I_2$	$I_1I_1I_2$
$I_2I_1I_2I_1$	$I_2I_1I_1$
$I_2I_1I_2I_1$	$I_2I_1I_1$
$I_2I_1I_1I_1$	$I_2I_1I_1$
$I_1I_1I_1\cdots$	$I_1I_1\cdots$

It is clear that varying numbers of commands will be processed, and that an improvement in efficiency has resulted from the addition of the store.

The expected efficiency, maximum expected efficiency, and average efficiency

$$E(a, n, \Phi, R), \quad EMAX(a, n, \Phi, R), \quad F(a, n, \Phi, R) \tag{6}$$

are defined as for the first model. The expected equipment-use efficiency $EU(a, n, \Phi, R)$ might be defined as

$$EU(a, n, \Phi, R) = \frac{\text{expected number of program steps processed}}{\text{a measure of equipment}}$$

$$= E\ \frac{\text{program step processing equipment}}{\text{a measure of equipment}}. \tag{7}$$

DETERMINATION OF PROBABILITY DISTRIBUTION, SEQUENTIAL MODEL

Let the model be defined by $a = (a_1, \ldots, a_m)$. As the set of instructions is scanned during each step, there will be a first instruction which cannot be processed during the next step, and which stops the scan. Let this instruction be one of the members of I_i. Since it cannot be processed, it must be the $(a_i + 1)^{th}$ I_i in the current string of instructions, and a_i I_i's will be processed. In addition, this I_i will be the first instruction processed in the second following step. In the worst case only a_i instructions will be processed in the next step (the current string of instructions begins with a_i+1 I_i's). In the best case, all $||a||$ processing elements will be used. The set composed of all sets of instruction strings which can be processed leaving a certain I_i to be processed first in the next time step is finite, and its members can be enumerated by an index, j. It is then clear that the program steps will be processed in one of the following sets

$$S_{ji} = (a_1, \ldots a_i, \ldots, a_m)j,$$

i being the instruction type for which $a_i = a_i$, and for which another instruction of type i has occurred. This stops the scan and insures that the next scan begins with an instruction of type i.

Let $Pr(S_{ji})$ be the probability that S_{ji} will be processed. Let $Pr(I_k:S_{ji})$ be the probability that S_{ji} will be able to be processed after the scan, given that the first command scanned is I_k.

Now

$$Pr(S_{ji}) = \sum_{k=1}^{m} \left[\sum_{j} Pr(S_{jk}) \right] Pr(I_k:S_{ji}) \tag{8}$$

for $i = 1, \ldots, m$; j as defined previously.

Write

$$PQ_k = \sum_{\text{all } j \text{ for each } k} Pr(S_{jk})$$

and sum the above equations over the correct set of j's for each i.

$$PQ_i = \sum_{k=1}^{m} \left[PQ_k \sum_{j_i} Pr(I_k:S_{ji}) \right], \quad i=1, \ldots m. \tag{9}$$

In addition

$$\sum_{k=1}^{m} PQ_k = 1. \tag{10}$$

The $Pr(I_k:S_{ji})$ may be determined for particular values of $Pr(I_i)$. The m+1 equations (9) and (10) have rank m and may be solved for PQ_k, $k=1, \ldots m$. The $Pr(S_{ji})$ may be found by direct substitution in (8). Clearly

$$Pr(a_1, a_2, \ldots a_m) = \sum_{i} Pr(S_{ji})$$

for the i and j corresponding to the set $(a_1, \ldots a_m)$.

$E(\alpha, \Phi, R)$, $EMAX(\alpha, \Phi, R)$, and $F(\alpha, \Phi, R)$ may be computed.

Consider the special case in which m=2 and the conditional distribution of I_1, I_2 is binomial (independent) in the instruction scan, I_1, I_2 occur with the probabilities a and b=1-a respectively, and the model is defined by (k, j).

Write S_{ji} as k, s, 1 or t, j, 2 respectively. Defining $a^{-1}=b^{-1}=0$, and writing C(s,t) for the binomial coefficients, Eq. (8) becomes

$$Pr(k, s, 1) = C(k+s-1, s)a^k b^s \sum_{i=0}^{j} Pr(k, i, 1) \qquad (11)$$

$$+ C(k+s-1, k)a^{k+1} b^{s-1} \sum_{i=0}^{k} Pr(i, j, 2) \qquad 0 \leq s \leq j.$$

$$Pr(t, j, 2) = C(t+j-1, j)a^{t-1} b^{j+1} \sum_{i=0}^{j} Pr(k, i, 1) \qquad (12)$$

$$+ C(t+j-1, t)a^t b^j \sum_{i=0}^{k} Pr(i, j, 2) \qquad 0 \leq t \leq k.$$

Writing $PI_1 = \sum_{i=0}^{j} Pr(k, i, 1)$, $PI_2 = \sum_{i=0}^{k} Pr(i, j, 2)$,

Eq. (10), (11), and (12) become

$$PI_1 = PI_1 a^k \left[\sum_{s=0}^{j} C(k+s-1, s) b^s\right] + PI_2 a^{k+1} \left[\sum_{s=0}^{j} C(k+s-1, k) b^{s-1}\right]$$

$$PI_2 = PI_1 b^{j+1} \left[\sum_{t=0}^{k} C(j+t-1, j) a^{t-1}\right] + PI_2 b^j \left[\sum_{t=0}^{k} C(t+j-1, t) a^t\right]$$

$$PI_1 + PI_2 = 1.$$

If

$$\Delta = 1 + a^{k+1} \sum_{s=0}^{j} C(k+s-1, k) b^{s-1} - a^k \sum_{s=0}^{j} C(k+s-1, s) b^s,$$

the solution may be written

$$Pr(k, s, 1) = \left(\frac{1}{\Delta}\right)\left[a^{2k+1} b^s C(k+s-1, s) \sum_{s=0}^{j} C(k+s-1, k) b^{s-1}\right.$$

$$\left. + a^{k+1} b^{s-1} \left(1 - a^k \sum_{s=0}^{j} C(k+s-1, s) b^s\right) C(k+s-1, k)\right] \quad 0 \leq s \leq j$$

$$Pr(t, j, 2) = \left(\frac{1}{\Delta}\right)\left[a^{k+t} b^{j+1} C(t+j-1, j) \sum_{s=0}^{j} C(k+s-1, k) b^{s-1}\right.$$

$$\left. + a^t b^j C(t+j-1, t) \left(1 - a^k \sum_{s=0}^{j} C(k+j-1, s) b^s\right)\right] \quad 0 \leq t \leq k.$$

Specializing still further, remembering

$$Pr(s, t) = Pr(s, t, 1) + Pr(s, t, 2),$$

let the model be defined by $(1_1, 1_2)$ or k=j=1,

$$P(1,0) = \frac{a^3}{a^2+b^2},$$

$$P(1,1) = \frac{ab}{a^2+b^2},$$

$$P(0,1) = \frac{b^3}{a^2+b^2},$$

$$E(1,1,a,b) = \frac{1}{2}P(1,0) + P(1,1) + \frac{1}{2}P(0,1) = \frac{1}{2}\left(1 + \frac{ab}{a^2+b^2}\right),$$

$$EMAX\ (1,1) = E\left(1,1,\frac{1}{2},\frac{1}{2}\right) = \frac{3}{4}.$$

If the distribution function $\Phi(a)$ is rectangular,

$$d\Phi(a) = \begin{cases} 0 & t<0 \\ dt & 0 \leqslant t \leqslant 1 \\ 0 & t>1 \end{cases}.$$

$$F(1,1) = 0.643.$$

Let the processing ability be defined as the number of program steps processed simultaneously. The maximum processing ability of this configuration is 1.5. In this case, the commands I_1, I_2 are equally likely to occur and satisfy a binomial distribution. The average processing ability, with varied programs being run on the same equipment and $\Phi(I_1)$ being rectangular, is 1.286.

Consider the model (k, 1) with the same probability assumptions.

$$\Pr(k, 0) = \frac{a^{2k+1}}{1-(k+1)a^k b} .$$

$$\Pr(k, 1) = \frac{a^k(1-a^k)}{1-(k+1)a^k b} .$$

$$\Pr(0, 1) = \frac{b[1-(k+1)a^k+ka^{2k}]}{1-(k+1)a^k b} .$$

$$\Pr(s, 1) = \frac{ba^s[(1-a^k)(1+sa^k b)+ba^k(sa^k-k)]}{1-(k+1)a^k b} \qquad 1 \leq s \leq k-1 .$$

$$E(k, 1, a, b) = \left(\frac{a^k(1-a^k b)}{1-(k+1)a^k b}\right)$$

$$+ \frac{1}{(k+1)(1-(k+1)a^k b)} \left[\sum_{s=1}^{k-1} ba^s[(1-a^k)(1+sa^k b)+ba^k(sa^k-k)](s+1)\right.$$

$$\left. + b(1-(k+1)a^k + ka^{2k}) - a^{2k+1}\right] .$$

Even if $k = 2$, the formula is too complicated for analysis; therefore, machine computation was used.

COMPUTATIONAL RESULTS, SEQUENTIAL MODEL

A 704 program was written for the case in which m=2, the conditional distribution of I_1, I_2 in the scan is binomial, and $\Phi(I_1)$ is rectangular. The cases covered ranged over combinations of

$$1 \leq a_1 = j \leq a_2 = k \leq 18.$$

For each configuration (j, k), subscripts having been dropped for ease of notation, the solution was tabulated for $0 \leq a \leq 1$ in steps of 1/50. Then EMAX (j, k) was found by direct search, and F(j, k) by integration using Simpson's rule.

Figure 1 shows the average expected efficiency F(j, k) as a function of processor distribution. Figure 2 shows the maximum expected efficiency EMAX (j, k) as a function of processor distribution. The average expected efficiency F(j, k) has these properties:

1. decreases rapidly as the distribution of a_1, a_2 processors becomes unbalanced,
2. increases rapidly as (j, k) approaches
$$\left(\frac{a_1+a_2}{2} \quad \frac{a_1+a_2}{2} \right),$$
3. increases slowly as m increases for the distributions (mj, mk).

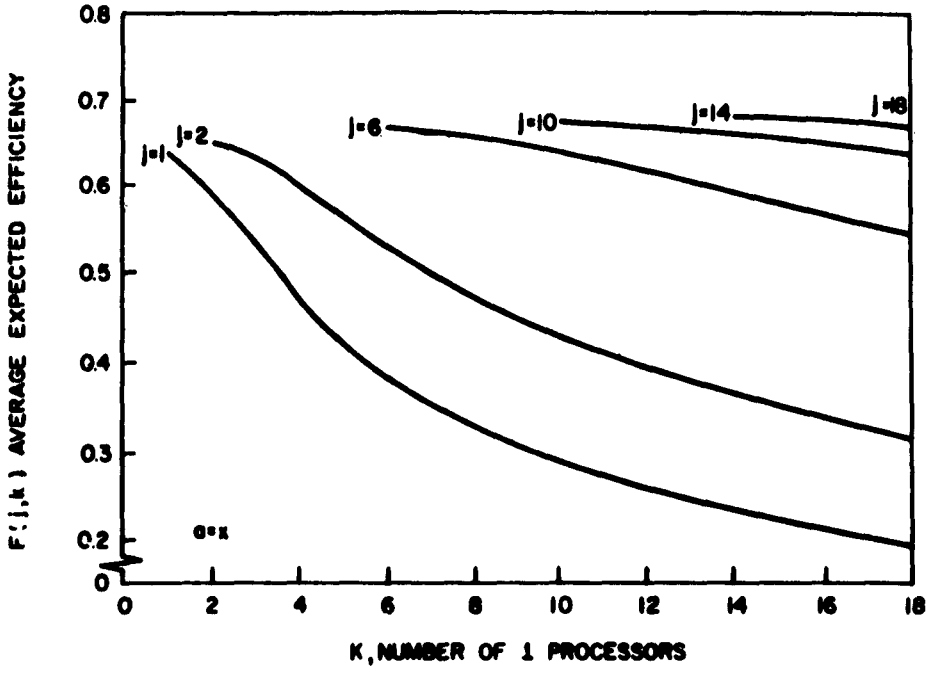


Figure 1 - Average expected efficiency as a function of processor distribution.

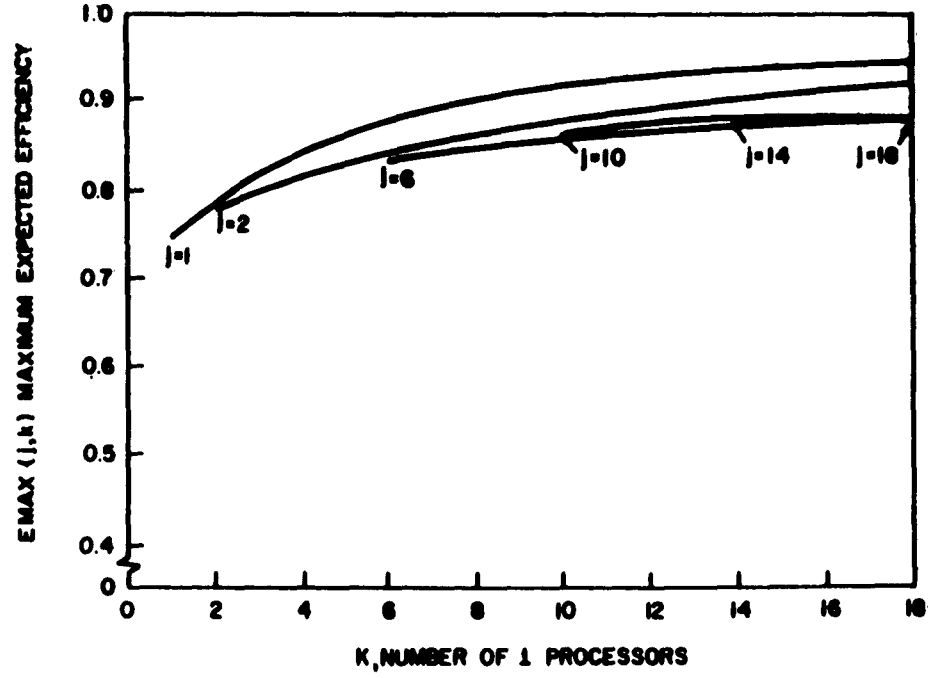


Figure 2 - Maximum expected efficiency as a function of processor distribution.

The maximum expected efficiency EMAX(j, k) has these properties:

1. increases slowly as number of processors increases,

2. increases as the distribution of a_1, a_2 becomes unbalanced.

The value $PR(I_2)$ for which EMAX(j, k) is assumed is greater than $\frac{k}{k+j}$.

The following table shows numerical results for important cases:

j	k	EMAX (j, k)	Corresponding Value $Pr(I_2)$	F (j, k)
1	1	.750	.5	.643
2	2	.804	.5	.661
3	3	.833	.5	.671
1	2	.798	.75	.601
1	3	.838	.825	.538
2	3	.824	.625	.648

These results show that

1. to achieve a high efficiency of computer operation during the running of a particular program, it is necessary to match the configuration of equipment with the probability of occurrence of program steps.

2. to achieve the greatest efficiency of computer operation during the running of a variety of programs, it is necessary to have a balanced distribution of equipment.

In most cases, achieving a high efficiency in running one program is done at the expense of overall efficiency of operation and vice versa. In addition, the efficiency of operation of an equipment configuration varies considerably as the probability of occurrence of program steps changes. To attempt to achieve greater efficiency in both cases, the second model, with store, will be analyzed.

DETERMINATION OF PROBABILITY DISTRIBUTION, SKIPPING SCAN

The behavior of the model is determined by:

1. the (m+1)- tuple $(a_1, a_2 \ldots a_m | n)$ giving the capacity (n) of the store , and the number and type (a_j) of processors $(n \geq \sum_{j=1}^{m} a_j)$,

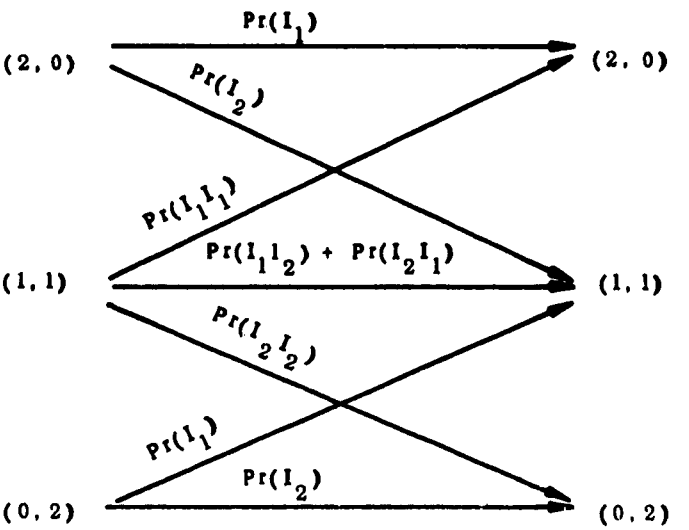
2. the distribution of commands I_j in the string which fills the store in the manner described previously.

Denote the contents of the store by $(a_1, \ldots a_m)$ where a_j denotes the number of instructions I_j in the store. Clearly

$$a_j \geq 0, \quad \sum_{j=1}^{m} a_j = n.$$

Let $Pr(a_1 \ldots a_m)$ denote the probability that the store is in the state $(a_1, \ldots a_m)$ at the beginning of a time step. During this time step, if $a_j \leq \alpha_j$, then a_j instructions I_j will be processed. If $a_j > \alpha_j$, α_j instructions I_j will be processed and the remaining $(a_j - \alpha_j)$ I_j instructions put into places of higher priority in the store so that they will be the first I_j instructions considered for processing during the next step. The empty places in the store will be filled by instructions as governed by their distribution.

The probability analysis proceeds by an analysis of storage state and is best explained by a simple example. Consider the model (1, 1| 2). The possible storage states are (2, 0), (1, 1), (0, 2). Applying the processing rules stated above, the storage state diagram,[1] with the probability of paths indicated on the paths, is:



This may be written in matrix form as

Input	$Pr(2,0)$	$Pr(1,1)$	$Pr(0,2)$	
	$Pr(I_1)$	$Pr(I_1I_1)$	0	$Pr(2,0)$
	$Pr(I_2)$	$Pr(I_1I_2)+Pr(I_2I_1)$	$Pr(I_1)$	$Pr(1,1)$
	0	$Pr(I_2I_2)$	$Pr(I_2)$	$Pr(0,2)$
				Output

where the elements represent the probability of proceeding from the input state to the output state as shown.

The equations whose solution gives $Pr(a_1, a_2)$ are thus seen to be:

$$\begin{pmatrix} Pr(I_1) & Pr(I_1I_1) & 0 \\ Pr(I_2) & Pr(I_1I_2)+Pr(I_2I_1) & Pr(I_1) \\ 0 & Pr(I_2I_2) & Pr(I_2) \end{pmatrix} \begin{pmatrix} Pr(2,0) \\ Pr(1,1) \\ Pr(0,2) \end{pmatrix} = \begin{pmatrix} Pr(2,0) \\ Pr(1,1) \\ Pr(0,2) \end{pmatrix} \tag{13}$$

$$Pr(2,0) + Pr(1,1) + Pr(0,2) = 1 \tag{14}$$

and may readily be solved.

Once the probabilities are found, the desired efficiencies follow immediately.

While the special case of models $(1, 1, n)$ can be solved explicitly for a binomial distribution of I_1, I_2, the general case is

more complicated. The n x n matrix of equations analogous to (13) may be readily generated column by column using the processing rules described earlier. The rank of this matrix is clearly at most n-1. In the special case, $Pr(I_1) = 1$, $Pr(I_j) = 0$, $j \neq 1$, the augmented matrix found from the analogs of (13) and (14) is seen to be nonzero. Thus, it may be assumed that a solution always exists and may be found numerically. A program has been written to generate and solve these equations in the case of some simple distributions of the I_j. Once generated, the equations are solved for $0 \leq a \leq 1$ at intervals of 1/50 and EMAX and F found as before. As usual, the expected efficiencies are calculated.

Now consider the model (1, 1, n) with a binomial distribution of I_1, I_2. For convenience, write $Pr(I_1) = a$, $Pr(I_2) = b$, $(b = 1-a)$. The set of equations to be solved may be easily found by the preceding rules. The matrix of coefficients is the following triple diagonal matrix, with the exception of the last row.

$$\begin{pmatrix} 1-a & -a^2 & 0 & & 0 \\ -b & 1-2ab & -a^2 & 0 & 0 \\ 0 & -b^2 & 1-2ab & -a^2 & 0 \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & -b^2 & 1-2ab & -a^2 \\ 0 & - & -b^2 & 1-2ab & -a \\ 1 & - & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} Pr(n,0) \\ Pr(n-1,1) \\ -- \\ -- \\ -- \\ -- \\ -- \\ Pr(0,n) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ - \\ - \\ - \\ - \\ 0 \\ 1 \end{pmatrix}$$

Using a+b=1, the determinant may be found to be

$$\Delta = \sum_{j=0}^{n-1} b^{2j} a^{2n-2j-2}$$

and the solution is

$$Pr(n,0) = a^{2n-1} \mid \Delta$$

$$Pr(0,n) = b^{2n-1} \mid \Delta$$

$$Pr(n-j,j) = (b^{2j-1} a^{2n-2j-1}) \mid \Delta \ .$$

$$E(1,1 \mid n,a,b) = \frac{1}{2} \quad (Pr(n,0)+Pr(0,n)) + \sum_{j=1}^{n-1} Pr(n-j,j)$$

$$= \frac{1}{2} + \frac{1}{2\Delta} \left(\sum_{j=1}^{n-1} b^{2j-1} a^{2n-2j+1} \right) .$$

Since $E(1,1|n,a,b)$ is a symmetric function, the maximum value occurs at $a = b = \frac{1}{2}$, so $EMAX(1,1|n) = 1 - \frac{1}{2n}$.
$F(1,1|m)$ can be obtained by numerical integration.

COMPUTATIONAL RESULTS - SKIPPING SCAN MODEL

A 704 program was written for the case of two processors, conditional binomial distribution for I_1 and I_2, rectangular distribution for $\Phi(I_1)$, and n instruction storage capacity. The cases covered ranged over

$$1 \leq a_1 = j \leq a_2 = k \leq 18$$

$$2 \leq j+k \leq n \leq 20.$$

The program scanned the input parameters, generated and solved the correct set of equations for $Pr(j,k|n)$ for each $Pr(I_1)$ and $Pr(I_2) = 1 - Pr(I_1)$. $Pr(I_1)$ ranged from 0 to 1 in 40 steps. Then $EMAX(j,k|n)$ and $F(j,k|n)$ were calculated using the appropriate formulas. Figure 3 shows $EMAX(j,k|n)$ and $F(j,k|n)$ for the cases covered above.

The average expected efficiency $F(j,k|n)$ has these properties:

1. it is greater than $F(j,k)$, all $n \geq j+k$,

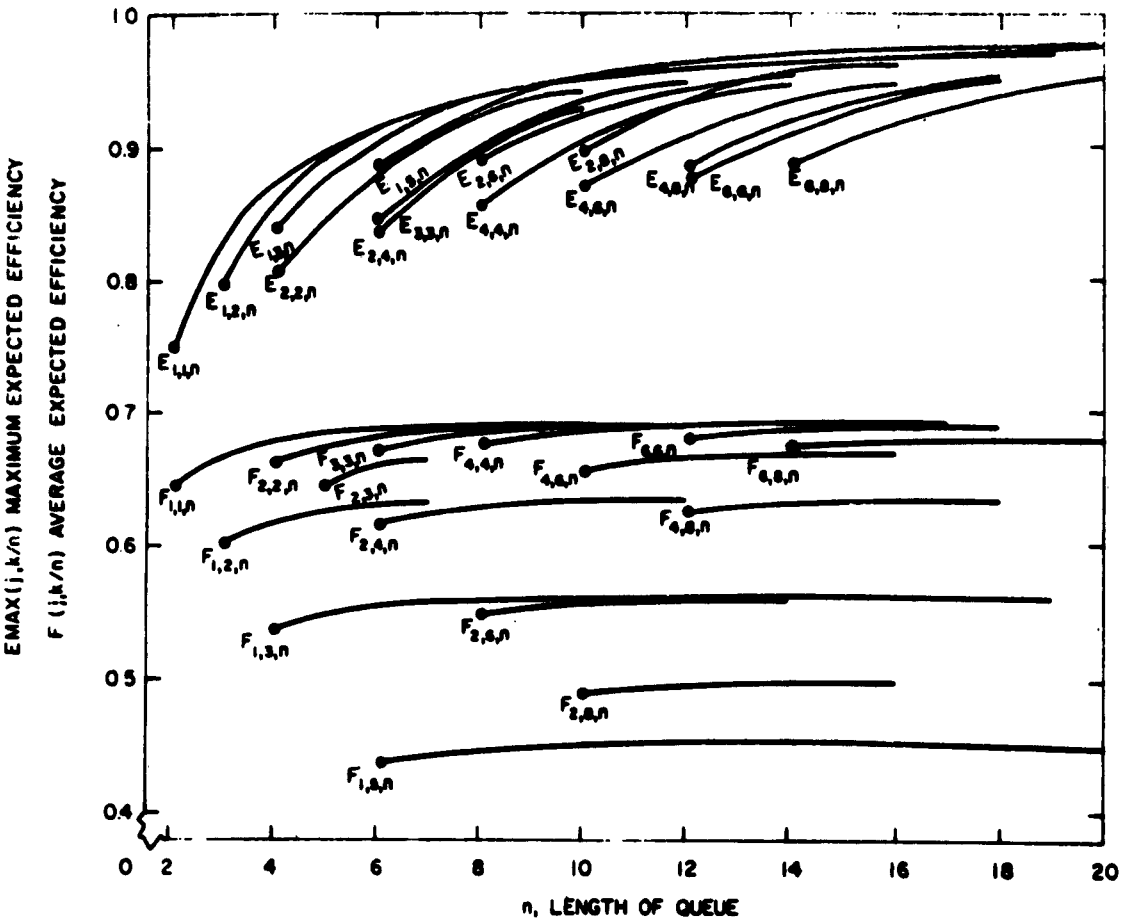


Figure 3 - Maximum and average expected efficiency as a function of queue length.

2. it increases slowly with the increase in store capacity,

3. it increases most slowly as distribution of a_1, a_2 processors becomes unbalanced,

4. it seems to approach distinct limiting values as m increases in the distributions (mj, mk).

The maximum expected efficiency EMAX(j, k|n) has these properties:

1. it is greater than EMAX(j, k), all $n \geq j+k$,

2. it increases rapidly with the initial increase in storage capacity as $\frac{dEMAX(1,1|2)}{dn} = \frac{1}{2n^2}$ and the other slopes approximate this,

3. it increases more slowly when the store is larger,

4. no matter what the initial distribution, it always approaches a value of about 95% or more.

These results show that:

1. the addition of small stores increases the efficiency of computer operation for programs in which the configuration of equipment and the probability of occurrence of program steps are well matched,

2. the addition of an auxiliary store aids slightly the efficiency of computer operation as a variety of programs is run,

3. the efficiency of operation of an equipment configuration varies as the probability of occurrence of program steps changes, but

less rapidly than without an auxiliary store.

MODEL MODIFICATIONS

The addition of store stations also uses equipment. Assume that a store uses $\frac{1}{x}$ as much equipment as a processing box. For the model (1, 1|n), the expected equipment-use efficiency

$$UE(1,1|n) = E(1,1|n) \quad \frac{\text{Processor equipment}}{\text{Total equipment}}$$

becomes

$$UE(1,1|n) = (1 - \frac{1}{2n})(\frac{x}{x+(n-2)})$$

clearly if

$2(n^2-n+1) > x$, UE (1, 1| n) decreases. The following table gives numerical results.

n	E(1, 1|n)	UE(1, 1|n) 1|x = 1|19	UE(1, 1|n) 1|x = 1|33
2	.75	.750	.750
3	.833	.791	.807
4	.875	.791	.825
5	.900	.760	.825

The increases in efficiency promised by adding auxiliary storage evidently depend strongly upon the extra equipment added.

Often the conditional distribution of the command in the scan is not binomial. Polya has proposed a distribution which seems probable.[2] If the initial probability of I_1 is a and of I_2 is b=1-a, then, if the first n trials resulted in k I_1's and (n-k) I_2's, the conditional probability of I_1, I_2 on the $(n+1)^{st}$ trial is

$$\frac{a + k\gamma}{1 + n\gamma}, \qquad \frac{b + (n-k)\gamma}{1 + n\gamma}, \qquad \gamma \geq -1 \text{ respectively.}$$

The probability of exactly k successes in the first n trials is given by $\pi(k, n) = C(n, k) \quad \frac{a(a+\gamma)\ldots(a+k\gamma - \gamma)\, b(+\gamma)\ldots(b+(n-k-1)\gamma)}{1(1+\gamma)(1+2\gamma)\ldots(1+(n-1)\gamma)}$.

If the model $(1, 1 \mid n)$ is analyzed,

$$E(1, 1 \mid n, a, b, \gamma) = 1 - \frac{1}{2} \quad \frac{a^n(a+\gamma)^{n-1} + b^n(b+\gamma)^{n-1}}{\Delta},$$

where $\Delta = a^n(a+\gamma)^{n-1} + b^n(b+\gamma)^{n-1} + ab \sum_{j=2}^{n} a^{n-j}(a+\gamma)^{n-j} b^{j-2}(b+\gamma)^{j-2}$.

$$EMAX(1, 1 \mid n, \gamma) = 1 - \frac{1 + 2\gamma}{2n + 4\gamma}.$$

Initially if $Pr(I_1) = Pr(I_2) = \frac{1}{2}$, then the following table shows the effect of γ.

γ	$Pr(I_1^2)$	$Pr(I_1I_2) = Pr(I_2I_1)$
0	.25	.25
1/2	.33	.167
1	.375	.125

The following table gives values of EMAX for $\gamma = 0$, $\frac{1}{2}$, and 1.

n	$E(1,1\|n,0)$	$E(1,1\|n,\frac{1}{2})$	$E(1,1\|n,1)$
2	.75	.667	.625
3	.833	.75	.7
4	.875	.8	.75
5	.9	.833	.786
10	.95	.911	.875

It is clear that the bunching of instructions of the same type decreases the efficiency of multiplexing.

CONCLUSIONS

The following statements based upon the previous computational results may be used as intuitive guides for the design of complex multiplexed equipment.

1. Matching the number and types of multiplexed processors used in the design to the expected probability of occurrence of instructions will

   a. give high peak efficiency if the match is right.

b. result in rapidly decreasing average efficiency if the combination of multiplexed processors is unbalanced.

2. Using balanced combinations of numbers and types of multiplexed processors in the design will give maximum average efficiency if combinations of instruction steps with different probabilities of occurrence are to be used.

3. The addition of stores to permit "skipping scans" will increase the peak efficiency of problem solving rapidly, especially for balanced combinations.

4. The increase in efficiency due to stores is very sensitive to the relative cost of storage stations.

5. The bunching of types of instruction steps will decrease multiplexing efficiency. "Skipping scans" decrease this effect.

ACKNOWLEDGMENT

This work was initiated after discussions with M. C. Andrews and W. L. Duda. The author, and this exposition, have benefited greatly from discussions with J. L. Smith and A. L. Leiner.

REFERENCES

1. P. M. Morse, "Queues, Inventories and Maintenance," (John Wiley & Sons, Inc., New York, 1958), 1st Edition, pp. 4-15.

2. W. Feller, "Probability Theory and Its Application," (John Wiley & Sons, Inc., New York, 1957), 2nd Edition., Vol. 1, p. 83.